

IAN F. SHAW
LA EVALUACIÓN
CUALITATIVA

INTRODUCCIÓN A LOS MÉTODOS
CUALITATIVOS

Paidós
Básica



La evaluación cualitativa

Paidós Básica

Últimos títulos publicados

63. R. Rorty - *Objetividad, relativismo y verdad*
64. R. Rorty - *Ensayos sobre Heidegger y otros pensadores contemporáneos*
65. D. D. Gilmore - *Hacerse hombre*
66. C. Geertz - *Conocimiento local*
67. A. Schütz - *La construcción significativa del mundo social*
69. M. Hammersley y P. Atkinson - *Etnografía*
70. C. Solís - *Razones e intereses*
71. H. T. Engelhardt - *Los fundamentos de la bioética*
72. E. Rabossi (comp.) - *Filosofía de la mente y ciencia cognitiva*
73. J. Derrida - *Dar (el) tiempo*
74. R. Nozick - *La naturaleza de la racionalidad*
75. B. Morris - *Introducción al estudio antropológico de la religión*
76. D. Dennett - *La conciencia explicada*
77. J. L. Nancy - *La experiencia de la libertad*
78. C. Geertz - *Tras los hechos*
79. R. R. Aramayo y otros - *El individuo y la historia*
80. M. Augé - *El sentido de los otros*
81. C. Taylor - *Argumentos filosóficos*
82. T. Luckmann - *Teoría de la acción social*
83. H. Jonas - *Técnica, medicina y ética*
84. K. J. Gergen - *Realidades y relaciones*
85. J. R. Searle - *La construcción de la realidad social*
86. M. Cruz (comp.) - *Tiempo de subjetividad*
87. C. Taylor - *Fuentes del yo*
88. T. Nagel - *Igualdad y parcialidad*
89. U. Beck - *La sociedad del riesgo*
91. K. R. Popper - *El mito del marco común*
92. M. Leenhardt - *Do Kamo. La persona y el mito en el mundo melanesio*
93. M. Godelier - *El enigma del don*
94. T. Eagleton - *Ideología*
95. M. Platts - *Realidades morales*
96. C. Solís - *Alta tensión: filosofía, sociología e historia de la ciencia*
97. J. Bestard - *Parentesco y modernidad*
98. J. Habermas - *La inclusión del otro*
99. J. Goody - *Representaciones y contradicciones*
100. M. Foucault - *Entre filosofía y literatura. Obras esenciales, vol. 1*
101. M. Foucault - *Estrategias de poder. Obras esenciales, vol. 2*
102. M. Foucault - *Estética, ética y hermenéutica. Obras esenciales, vol. 3*
103. K. R. Popper - *El mundo de Parménides*
104. R. Rorty - *Verdad y progreso*
105. C. Geertz - *Negara*
106. H. Blumenberg - *La legibilidad del mundo*
107. J. Derrida - *Dar la muerte*
108. P. Feyerabend - *La conquista de la abundancia*
109. B. Moore - *Pureza moral y persecución en la historia*
110. H. Arendt - *La vida del espíritu*
111. A. MacIntyre - *Animales racionales y dependientes*
112. A. Kuper - *Cultura*
113. J. Rawls - *Lecciones sobre la historia de la filosofía moral*
114. Th. S. Kuhn - *El camino desde la estructura*
115. W. V. O. Quine - *Desde un punto de vista lógico*
116. H. Blumenberg - *Trabajo sobre el mito*
117. J. Elster - *Alquimias de la mente*
118. I. F. Shaw - *La evaluación cualitativa*

Ian F. Shaw

La evaluación cualitativa

Introducción a los métodos cualitativos

Título original: *Qualitative Evaluation*
Originalmente publicado en inglés, en 1999, por Sage Publications, Londres,
Thousand Oaks y Nueva Delhi

Traducción de Roc Filella Escolà

Cubierta de Mario Eskenazi

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del *copyright*, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

© 1999 Ian Shaw
© 2003 de la traducción, Roc Filella Escolà
© de todas las ediciones en castellano
Ediciones Paidós Ibérica, S. A.,
Mariano Cubí, 92 - 08021 Barcelona
y Editorial Paidós, SAICF,
Defensa, 599 - Buenos Aires
<http://www.paidos.com>

ISBN: 84-493-1426-7

Depósito legal: B-20.354/2003

Impreso en A & M Gràfic, S.L.
08130 Santa Perpètua de Mogoda (Barcelona)

Impreso en España - Printed in Spain

A mis padres

SUMARIO

Agradecimientos	13
1. Al encuentro de la evaluación cualitativa	15
Giros equivocados, puestos fronterizos y fuego amigo	16
Tendencias de la evaluación cualitativa	20
La <i>evaluación</i> cualitativa	24
La evaluación <i>cualitativa</i>	29
Las características de la indagación cualitativa	30
Sinopsis	34
2. Teóricos de la evaluación y evaluación cualitativa: de	
persuaciones y persuasivos	39
Donald Campbell: investigar las causas	41
Michael Scriven: una ciencia de la evaluación	43
Lee Cronbach: evaluar dentro de los programas	45
Robert Stake: la evaluación receptiva	48
Carol Weiss: evaluación de políticas	50
Ernest House: evaluar para la justicia social	54
La evaluación crítica	57
William Reid: evaluación para desarrollar la práctica	59
Elliot Eisner: la evaluación reflexiva	62
Egon Guba e Yvonna Lincoln: la evaluación constructivista	65
Conclusión: agenda preliminar para la evaluación cualitativa	67
3. Las pruebas obtenidas con la evaluación cualitativa	71
Paradigmas y pragmatismo	73
El pospositivismo realista	75
La evaluación crítica	77
El constructivismo	79
El pragmatismo	81
<i>Ejemplo 3.1. ¿Puede ser radical el pragmatismo?</i>	85
Realismo y relativismo	86
¿Rigor o relevancia?	92
4. Valores, validez y los usos de la evaluación	99
Validez y valores	99
Verosimilitud y credibilidad	101

<i>Ejemplo 4.1. La idea de validez de Hammersley</i>	101
La evaluación como actividad moral y práctica	104
La evaluación reformista	106
Los criterios abiertamente ideológicos	107
Cuestiones sobre la validez	109
¿Ingeniería o ilustración?	112
La ética de la evaluación	114
<i>Ejemplo 4.2. La ética: ¿ha fracasado la ciencia positivista?</i>	115
Metodología y métodos	117
Conclusión	120
5. Evaluar programas y políticas	123
La evaluación de programas	124
<i>Ejemplo 5.1. La evaluación de programas prenatales</i>	126
Temas para los evaluadores cualitativos de programas	128
Procesos, resultados y racionalidad	129
Teorizar sobre los programas	131
<i>Ejemplo 5.2. Teorías críticas de los programas</i>	132
Evaluación y rendición de cuentas	133
Evaluación y cambio social	135
Evaluación de políticas y uso de la información	136
Evaluación de políticas y política social	142
6. La evaluación del profesional	147
Preguntas para la indagación del profesional	148
La indagación y la evaluación del profesional	149
Conocimientos tácitos e indagación cualitativa	151
La investigación del profesional	152
El profesor como investigador	153
La investigación clínica	155
La investigación cualitativa del profesional de la asisten- cia social	156
La investigación crítica del profesional	158
Críticas etnográficas	159
La evaluación participativa	161
La indagación participativa	162
Participación en la evaluación de defensa	165
Evaluar en la práctica	168
Las pruebas empíricas	170
La indagación reflexiva	171
La evaluación cualitativa para los profesionales	173
7. Las decisiones sobre el diseño de la evaluación	177
Valorar la evaluabilidad	180

<i>Ejemplo 7.1. La valoración de la evaluabilidad</i>	182
Ética del diseño y justicia social	183
<i>Ejemplo 7.2. El consentimiento informado</i>	184
Evaluar para la justicia social	186
Diseñar la evaluación cualitativa para los resultados	188
Campbell y el estudio de casos	189
Los microprocesos	191
Causa y modelos causales	192
Los estudios de casos en la evaluación	194
<i>Ejemplo 7.3. El estudio en casos de enclaves múltiples</i> . . .	198
<i>Ejemplo 7.4. El estudio de casos para los profesionales y los investigadores</i>	200
La acción y la participación	201
<i>Ejemplo 7.5. La ciencia educativa crítica</i>	202
8. La evaluación en el campo	205
Evaluar los métodos de trabajo de campo	207
<i>Ejemplo 8.1. Analizar la toma de decisiones del profesional</i>	207
La entrevista evaluativa	209
<i>Ejemplo 8.2. La revisión cultural</i>	211
La observación	215
Las historias de la vida	216
<i>Ejemplo 8.3. Problemas con la vivienda</i>	218
Textos y documentos	219
Los grupos centrados	221
Las simulaciones	223
Temas y cuestiones del trabajo de campo de la evaluación cualitativa	225
El acceso	225
Evaluación y temas delicados	227
La satisfacción	228
La evaluación de potenciación y del profesional	231
<i>Ejemplo 8.4. Autoevaluación de la terapia familiar con niños que sufren trastornos de conducta</i>	231
<i>Ejemplo 8.5. Profesores que colaboran en la potenciación</i>	232
La ética del trabajo de campo	234
<i>Ejemplo 8.6. Ética e investigación de la indigencia</i>	236
9. El análisis: establecer las afirmaciones de la evaluación	239
Las afirmaciones de la evaluación	240
Las justificaciones cualitativas	242
El análisis de las entrevistas de la evaluación cualitativa	243
<i>Ejemplo 9.1. La codificación indígena emergente</i>	245

<i>Ejemplo 9.2. Fomentar los vínculos narrativos visibles . . .</i>	246
El análisis participativo y del profesional	247
Ética y análisis	250
<i>Ejemplo 9.3. Grupos lúdicos infantiles: ¿de qué lado estamos?</i>	252
Las causas y las explicaciones	252
<i>Ejemplo 9.4. Mapas causales: modelos del participante para un centro de actividades rural</i>	254
La validación de los miembros y la triangulación	256
Las síntesis y el metaanálisis	260
La generalización	261
La lógica de la encuesta	263
La generalización analítica	263
Transferencia caso por caso y «descripción gruesa»	264
Dar relevancia a la evaluación	265
¿Coautoría?	266
<i>Ejemplo 9.5. Los principios democráticos en la práctica . .</i>	267
Presentación de las conclusiones	268
<i>Ejemplo 9.6. Lee Cronbach sobre las historias contables . .</i>	269
La defensa	270
Apéndice: ejercicios	273
Bibliografía	289
Índice de nombres	307
Índice analítico	313

AGRADECIMIENTOS

Gracias a Jonathan Scourfield, Nicki Barnes, Paula Harris, Colin Lewis y, una vez más, a Nick Evans por suavizar algunas de las exigencias cotidianas de la vida, de una forma que influyó mucho en la redacción de este libro. Y a Anthea, por la caravana.

CAPÍTULO 1

AL ENCUENTRO DE LA EVALUACIÓN CUALITATIVA

Lo que necesita el campo de la evaluación es un buen antropólogo social.

LEE CRONBACH

Han pasado ya más de treinta años desde que Robert Stake oyó cómo Lee Cronbach hacía el comentario anterior. Lo que a principios de los años sesenta era clarividente y radical, en este cambio de milenio resulta de una contundencia y una oportunidad desconcertantes. En todos los ámbitos de la vida pública, desde las cárceles a la farmacia, desde la política a la asistencia social, se oye la aclamación triunfalista de los éxitos de la acción basada en pruebas. Se cree que las evaluaciones controladas de las intervenciones sociales, medidas por los resultados conductistas, prometen poder verificar qué es lo que funciona y qué lo que no funciona. Si los profesionales no saben leer los resultados, ver las implicaciones obvias y actuar en consecuencia, mucho peor para ellos. Pero, a decir verdad, lo habitual es que las mediciones de la responsabilidad cuantitativas y orientadas a los resultados carezcan de la intensidad, la sutileza, la particularidad, el juicio ético y la relevancia que potencialmente caracterizan a la evaluación nacida de la etnografía.

Como contrapunto deliberado, la aparición de este libro como parte de una serie de introducción a la metodología cualitativa reafirma que la relación entre los propósitos evaluativos y la ciencia social cualitativa es un —posiblemente *el*— punto de apoyo para generar una buena investigación de evaluación.

Escribir este libro ha sido una experiencia desconcertante, un camino constante, en sus primeras fases, desde el «conocimiento» a la ignorancia. Tal vez así deba ser. Pensaba que sabía algo sobre la evaluación. Después de todo, la había practicado y había escrito sobre el tema. Pero pronto descubrí que, pese a todos los intentos y todos los

propósitos, nunca antes había reflexionado sobre ella. No partí de las temáticas «tendencias de la evaluación cualitativa», que se dilucidan más adelante en este capítulo introductorio. Acabé con ellas, como persuasiones.

GIROS EQUIVOCADOS, PUESTOS FRONTERIZOS Y FUEGO AMIGO

La evaluación contiene retos, temas e innovaciones apasionantes. Pero yo pecaría de ceguera ingenua si diera la impresión de que todo lo que se necesita es aplicar a la evaluación la metodología cualitativa, para que todo se ilumine y, quizá, se endulce también. Existen unas cuantas ambigüedades importantes tanto sobre el sujeto como sobre el objeto de esa aplicación: sobre la influencia de la clase dirigente de la evaluación, los términos en que se desarrollan los debates, la relación entre evaluación y ciencia social, la posición de la evaluación cualitativa y la reacción de la etnografía de carácter general ante la empresa evaluadora. Analizaré todas estas cuestiones con mayor detalle, como telón de fondo para la posterior explicación de los puntos en los que se insiste en los capítulos siguientes y que son su signo distintivo.

La clase dirigente de la evaluación sitúa persistentemente la evaluación cualitativa, en el mejor de los casos, como una parte de una solución de «caballo según carrera» a la metodología de la evaluación y, en el peor de los casos, como el Morris Minor de la familia Rolls Royce de la evaluación. Pocos gustan de levantar la mano para admitir que pertenecen a la clase dirigente. Pero no hay duda de que existe tal clase en el mundo de la evaluación. En ella se incluyen quienes se amparan bajo el manto de la primera generación de teóricos de la evaluación, producen unos textos respetados sobre los fundamentos de la evaluación de programas, configuran los planes evaluadores de los gobiernos o llegan a presidentes de la American Evaluation Association. Quienes proponen el enfoque del «caballo según carrera» de la evaluación creen, para seguir con la imagen, que, cuando el suelo de la pista está duro o firme, el caballo puede ir montado con toda seguridad por los diseños de una evaluación orientada a los resultados, mientras que, cuando el suelo está blando, hay que optar por una montura que sepa sortear sin peligro la comprensión de los difíciles procesos institucionales.

Sin embargo, sigue viva la acechante sospecha de que, en un mundo de días soleados y de secas carreras de evaluación, serán los mé-

todos de elección los que nos prometerán una información sólida sobre los resultados, los efectos y la eficiencia. La evaluación cualitativa servirá cuando el suelo esté blando, pero, tarde o temprano, éste se secará y será posible y deseable averiguar si los programas, las políticas y los proyectos han producido algún efecto apreciable. Para ese tipo de preguntas, el juicio clínico aleatorio y controlado ofrece el «criterio de oro, el Rolls Royce de los planteamientos de la evaluación» (Chelimsky, 1997, pág. 101). Lamentablemente, ni el elitismo de la cuasi-experimentación ni el pluralismo ecléctico del «caballo según carrera», con su supuesto tácito de «vive y deja vivir», pueden proporcionar una evaluación relevante y rigurosa.

En el mundo cotidiano de la práctica de la evaluación siguen rigiendo la inquietud por las mediciones, las preocupaciones tradicionales por la fiabilidad y la validez, y otros intereses que encierra la metodología cuantitativa. Estas preocupaciones han cobrado una nueva vitalidad y un nuevo sentido al paso del entusiasmo por los planteamientos basados en pruebas de toda una variedad de intervenciones sociales y de su evaluación. Tanto en ámbitos más generales como en otros más particulares existen versiones de la práctica basada en pruebas. La versión de ámbito general se rige sobre todo por la práctica. Incluye un énfasis en la responsabilidad, unas estrechas relaciones de trabajo entre investigadores y profesionales, una obligación general de basar en pruebas los juicios sobre la práctica, y un recurso explícito a las conclusiones de las investigaciones para apoyar las decisiones. Incluye también un programa que facilite el acceso a los datos y favorezca la difusión de los descubrimientos de las investigaciones de forma inteligible, una orientación hacia los resultados de la práctica, y el acento puesto en un desarrollo profesional factible e incremental.

La versión más centrada y particular de la práctica basada en pruebas suele regirse por criterios académicos. Este uso más restringido del término incluye las características de la versión más general, pero habitualmente conlleva además una confianza en la medición, la instrumentación, la objetividad de las pruebas, unos índices conductistas de buena práctica profesional, la cuasi-experimentación y, en algunos casos, una actitud negativa ante los métodos de la evaluación cualitativa.

Mis reservas se refieren sobre todo a la versión de ámbito más restringido. La propia expresión en cierto modo se delata. ¿Qué entendemos por «pruebas»? ¿Qué entendemos por «práctica»? ¿Cómo se «basa» una en las otras? Se dan por supuestas demasiadas cosas. Pe-

ro, evidentemente, uno no describe el santo grial, sino que se limita a buscar.

La confianza que se tiene en la medición en el ámbito de la práctica basada en pruebas delata la ingenua idea de que es muy fácil conseguir una comprensión de la realidad. El fracaso en cuestionar la relación existente entre las pruebas y la práctica es el talón de Aquiles de la práctica basada en pruebas. Muchos de los autores que comparten esta convicción dan por supuesto que la utilización de los resultados de la evaluación es algo relativamente directo. Cuando existe un problema, se atribuye a un fallo de comunicación o a la resistencia interesada de los profesionales (Sheldon y Macdonald, 1989-1990). Abordar así los problemas de la teoría y la práctica supone relegar la «práctica» al subordinado, a aquel sobre quien se actúa, y no reconocer que los problemas prácticos de este tipo se producen en el transcurso de *cualquier* actuación. Presumir que esta clase de dificultades de algún modo se pueden identificar y tratar en teoría, para luego aplicarlas en la práctica, suele ocultar el hecho de que, en realidad, surgen de la práctica. Así pues, la práctica basada en pruebas hace una torpe interpretación de cómo se debería utilizar la investigación. Para bien o para mal, está emparejada con los rígidos modelos de responsabilidad de la evaluación, y sus abogados suelen ser incapaces de plantear en serio los temas del poder. Casi todas las formas de la práctica basada en pruebas son no participativas. Las cuestiones de género y de raza se suelen tratar como problemas técnicos que se pueden mitigar, si no resolver por completo. Desgraciadamente, unos diseños bien ajustados no llevan a la verdad, y «el tecnicismo incuestionado en la investigación de evaluación es peor que la completa inexistencia de evaluación» (Palumbo, 1987, pág. 31).

Si la reticencia a abandonar la confianza en soluciones de diseño técnico es el primer gran problema, el segundo surge de unos intentos equivocados de delimitar para la evaluación un territorio que se aleje de la metodología general de las ciencias sociales. Veremos que en la mayoría de los temas es difícil encontrar teóricos de la evaluación más distintos entre sí que Michael Scriven, por un lado, e Yvonna Lincoln y Egon Guba, por el otro. Y sin embargo, coinciden en creer sin reservas que la investigación y la evaluación son empresas diferentes, y que decir lo contrario es un error y un serio perjuicio. Más adelante, en este mismo capítulo, volveremos al tema de la relación establecida entre evaluación e investigación. Aquí me limito a afirmar que considero que esta demanda de divorcio, o al menos de separación, es una mala noticia para la evaluación. Quizás esté equi-

vocado, pero me pregunto si su origen no estará en parte en un etnocentrismo norteamericano que propugna la identidad ocupacional separada de los evaluadores.

Se dan por supuestas otras fronteras que limitan y constriñen artificialmente el desarrollo de la evaluación. El hecho de que la evaluación se base predominantemente en la obra de los teóricos y los investigadores de la educación ha sido, a la vez, su virtud y su defecto. Las ventajas de compartir un vocabulario y unas preocupaciones prácticas del mundo educativo han sido la base de algunas de las mejores obras sobre la evaluación, especialmente sobre la evaluación cualitativa. Pero tal hecho produce miopía y una estrechez de miras que hace ya mucho que no tiene sentido. Es triste que el problema apenas se reconozca, y todavía menos se aborde. Cronbach es casi el único teórico de una perspectiva educativa que ha identificado y lamentado el daño que produce la ausencia casi completa de un trabajo que traspase fronteras entre quienes se plantean problemas relacionados con la evaluación en los campos de la salud, el derecho, la asistencia social y la educación. El etnocentrismo profesional es tan debilitante y potencialmente tan discriminatorio como el etnocentrismo nacional.

Existe otro puesto fronterizo que se debe dismantelar. En muchos de los escritos sobre la evaluación, cuando se dice «evaluación» se puede leer «evaluación de programas». Teorizar sobre la evaluación, la ubicación de ésta dentro de los procesos de cambio político y social, y los debates sobre cómo se utiliza y cómo se debería utilizar la evaluación, todo se suele considerar como problemas de comprensión y evaluación de *programas* de relativamente gran escala. La obra de Carol Weiss sobre evaluación de *políticas* es muy conocida, pero con demasiada frecuencia se ignora en la literatura sobre la evaluación. Asimismo, se habla mucho de la obra de teóricos como Robert Stake y Elliot Eisner, pero se comprenden muy poco las consecuencias que tiene para la evaluación de la *práctica* directa y local. La hegemonía de la evaluación de programas se ha diluido un poco. Por ejemplo, la American Evaluation Association, sobre todo por la influencia de Scriven, tiene unos criterios distintos para la evaluación de programas y la evaluación de personal. Pero esto, aunque pueda avalar la pretensión de un territorio ocupacional, poco sirve para fomentar los intereses que son más esenciales para la evaluación cualitativa.

La evaluación se encuentra ante otros dos retos importantes que no se han afrontado aún. Uno de los avances más significativos de los últimos diez años ha sido la progresiva aparición de unas formas de

investigación y de evaluación abiertamente ideológicas. Gran parte de la teoría de la evaluación, incluida la que, con sentido práctico, reconoce el carácter esencial de los temas políticos, es anterior a ese avance. En consecuencia, la clase dirigente de la evaluación ha establecido una jerarquía y ha situado, en la postura mayoritaria dentro de la práctica de la evaluación, las bases para excluir la evaluación de defensa o propugnación. Chelimsky representa esta postura. Confiesa la autora que la evaluación necesita realmente unos criterios de calidad y de práctica que tengan sentido en todo el campo, y no sólo en uno de sus segmentos: «Será fundamental que los criterios sean integradores de todos los legítimos propósitos de la evaluación, y no se limiten a santificar una perspectiva a expensas de otra» (Chelimsky, 1997, pág. 108). Pero concluye que «la honradez es esencial y la defensa no es realmente una opción» (pág. 109) y probablemente no se puede incorporar en los criterios comunes. El evaluador abiertamente ideológico dirá, por supuesto, que esto supone interpretar mal el problema y que los criterios generales no son una opción.

El último problema de la evaluación cualitativa tiene unos orígenes menos esperados. La etnografía de carácter general y algunos intereses importantes dentro de la sociología británica han mostrado un completo desdén hacia la evaluación cualitativa. Los orígenes de este fenómeno se encuentran en el desarrollo de las disciplinas de las ciencias sociales (Finch, 1986). Pero cualquiera que sea la razón, el fuego amigo ha sido uno de los factores más importantes en el debilitamiento de la evaluación cualitativa y en el fomento de nuevas fronteras artificiales entre la investigación y la evaluación cualitativas.

TENDENCIAS DE LA EVALUACIÓN CUALITATIVA

Mi intención en este libro es defender una evaluación cualitativa que responda a todos estos retos, estas limitaciones y estos giros equivocados. Los *motivos* rectores de la obra se pueden expresar como una serie de proposiciones:

- La evaluación se distingue por un grupo de propósitos evaluativos. Así se podrá dar respuesta verosímil y productiva a las cuestiones de elección de métodos, teoría y ética de la evaluación y evaluación de defensa o propugnación.
- La mejor forma de entender la evaluación es suponiendo que conlleva la realización de una *investigación* evaluativa, más que como

una serie diferenciada de axiomas o de métodos de evaluación, separada del trabajo de investigación más amplio.

- Los retos y las limitaciones particulares con que se enfrenta la evaluación normalmente exigirán optar por una evaluación cualitativa como metodología.
- Una evaluación relevante y rigurosa exige el desarrollo de una teorización y unas estrategias de evaluación de tipo interprofesional.
- La teoría y la metodología de la evaluación deben tanto a teóricos de la evaluación confesos, como a la obra de autores e investigadores que no se considerarían como tales.
- La evaluación cualitativa promete unas perspectivas diferenciadas pero coherentes de la evaluación de políticas, de programas y de la práctica.
- La evaluación cualitativa ofrece unas soluciones parciales creíbles a los problemas de la evaluación de análisis causal y de resultados.
- La evaluación cualitativa puede evitar la incapacidad sentimental de poner en entredicho sus propios análisis y soluciones.

La explicación de estas proposiciones está en todo el conjunto de lo que sigue. De momento sólo quiero subrayar la primera y la última. La defensa del *propósito evaluativo* es quizás el tema más omnipresente en este libro: «Una cierta idea de propósito es, en su raíz, lo que guía a cualquier práctica de la evaluación» (Greene, 1994, pág. 539). El tema, como tal, no es especialmente novedoso. Silverman dice de la investigación cualitativa en general que «las múltiples lógicas de la investigación cualitativa surgen de sus relaciones con los propósitos generales de los proyectos de investigación» (1997b, pág. 25).

No existe un único propósito evaluativo, sino muchos. Entre ellos están los planes de evaluación tradicionales de rendición de cuentas sobre los resultados de programas y políticas de carácter público, y de determinación de su eficacia. Pero entre los propósitos evaluativos está también el de llegar a comprender los problemas públicos y los esfuerzos pasados y presentes por abordarlos. En esto se incluirán las explicaciones normativas desde el punto de vista de la injusticia. La evaluación también pretende entender cómo funcionan y cómo cambian las organizaciones, y puede desarrollar y valorar unos medios para fortalecer las instituciones y mejorar el rendimiento. También son objetivos fundamentales de la evaluación una creciente receptividad como instrumento agente ante el público y los usuarios

de los servicios, y trabajar por reformar el gobierno mediante el libre flujo de información sobre la evaluación (Chelimsky, 1997).

Chelimsky vincula esta diversidad de propósitos a tres diferentes perspectivas conceptuales que de forma convincente señala dentro de la evaluación. Las describe como:

1. Evaluación para la *rendición de cuentas* (por ejemplo, medir los resultados o la eficiencia).
2. Evaluación para el *desarrollo* (por ejemplo, ofrecer una ayuda evaluativa para fortalecer las instituciones).
3. Evaluación para *conocer* (por ejemplo, conseguir un mejor conocimiento de alguna área específica o de un campo político).

No es necesario aceptar todas las implicaciones que según Chelimsky se derivan de este análisis. Por ejemplo, hemos visto que la autora defiende una visión de «caballo según carrera» de la tarea evaluadora. Con ello advierte contra el uso de estrategias cualitativas (por ejemplo, el estudio de casos) para abordar cuestiones y problemas que ella cree que deben seguir en el terreno de los métodos de orientación cuantitativa que buscan el control mediante el diseño. También excluye los métodos de defensa en aras de preservar los criterios y las ideas comunes de la calidad de la evaluación. Ninguna de estas posturas es útil. Pero la idea básica es valiosa, especialmente si, dentro de la diversidad, subrayamos la coherencia que la familia ampliada de los propósitos evaluativos permite: «La evaluación no es monolítica, pero nos hemos comportado como si lo fuera» (Chelimsky, 1997, pág. 104). Permítaseme que, a título ilustrativo, señale unos cuantos puntos en los que esta perspectiva resulta provechosa.

En primer lugar, evita el escollo que conlleva suponer la existencia de un único cuerpo de teoría de la evaluación que sirve de base para toda práctica evaluadora. A lo largo de este libro partimos de cuatro literaturas diferentes: la teoría de la evaluación, la evaluación educativa, la evaluación de la asistencia social y la metodología cualitativa de investigación. No me ocupo exhaustivamente de todos estos temas, pero sí, espero, de forma razonablemente completa. El alcance de mi exposición es la principal prueba que aporto de la proposición de que la teoría y la metodología de la evaluación deben tanto a teóricos de la evaluación confesos, como a la obra de autores e investigadores que no se considerarían como tales.

En segundo lugar, permite que las decisiones referentes a los métodos de evaluación sean a la vez amplias y particulares. En palabras

de Chelimsky, «aunque las tres perspectivas *pueden* usar cualquier método, dado que los propósitos son diferentes, lo habitual es que los métodos se empleen también de forma distinta» (1997, pág. 103). Habrá usos distintivos de los métodos entre los investigadores y los evaluadores convencionales, y también, como indica Chelimsky, dentro de la evaluación. Tendremos ocasión de reconocer e ilustrar este punto una y otra vez a lo largo de este libro. La ajustada articulación de la metodología y el propósito evita la tendencia, a menudo presente en el debate sobre la evaluación, de oponer la metodología a la ideología. La evaluación cualitativa que esté divorciada del propósito se convertirá en una simple reinención del tecnicismo del que tan a menudo ha sido culpable la evaluación cuantitativa. No deberíamos definir la evaluación cualitativa por los métodos. Por otro lado, la evaluación definida exclusivamente desde la perspectiva de la ideología nos deja sin estrategia y sin dirección a corto y medio plazo, y ante el peligro de nuestras propias pasiones.

En tercer lugar, la polémica sobre el propósito permite dejar «en suspenso» el difícil tema de la evaluación abiertamente ideológica, sin tener que recurrir ni a la excomunión ni a la beatificación.

La influencia de la última de mis proposiciones es también omnipresente en este libro. El *sentimentalismo* es un riesgo recurrente para todo aquel que se dedique a una forma u otra de indagación disciplinada (Shaw, 1999). Cuando se trata de la investigación y la evaluación, tanto los profesionales como los teóricos son vulnerables al sentimentalismo, que se distingue por la alimentación de los mitos y por una tendencia hacia prácticas supersticiosas.

Pecamos de sentimentalismo, dice Howard Becker,

cuando, por la razón que sea, rechazamos investigar algún asunto que debería considerarse problemático. Somos sentimentales, en especial, cuando nuestra razón es que preferiríamos no saber qué pasa, si saberlo supusiera violar alguna simpatía de cuya existencia es posible que ni siquiera seamos conscientes (1970, págs. 132-133).

Los sentimentalismos y las prácticas supersticiosas que me interesan de forma especial son aquellos que están estrechamente asociados con compromisos que yo mismo comparto. Si a lo largo de este libro hay de vez en cuando alguna nota demoledora, es porque así debe ser. Creo que era Filipo de Macedonia quien se servía de un hombre que llevaba un bastón en cuyo extremo superior había una vejiga de cerdo. La única función de ese hombre era ejercer la liber-

tad de despertar a Filippo a cualquier hora de la noche y golpearle la cabeza con la vejiga, para recordarle que era mortal. He intentado esparcir varias de esas vejigas a lo largo de este libro.

Hay dos preguntas que de momento han quedado en parte sin respuesta. La primera, cuáles son las diferencias entre investigación y evaluación, si es que existe alguna. La segunda, qué afirmamos cuando decimos que una evaluación es cualitativa. He dado mi respuesta a la primera pregunta, pero no he reconocido los contraargumentos a favor de una separación más radical entre investigación y evaluación. Nada he dicho sobre la segunda pregunta. Vamos a verlas a continuación.

LA EVALUACIÓN CUALITATIVA

Nadie, que yo sepa, querrá decir que los términos «evaluación» e «investigación» son distintos pero que no existe diferencia alguna que los distinga. Pero aquí es donde casi terminan los acuerdos. En la tabla 1.1 se recogen algunas de las distinciones que se han hecho entre evaluación e investigación.

Antes de comentar estas posibles distinciones entre evaluación e investigación, es necesario aclarar brevemente la cuestión. Al preguntar si la evaluación y la investigación son diferentes, planteamos tres preguntas distintas:

1. ¿La investigación y la evaluación *pueden* ser similares/diferentes?
2. ¿La investigación *es* diferente de la evaluación?
3. ¿La investigación y la evaluación *deberían* ser diferentes o similares?

La primera pregunta es teórica; la segunda, empírica; y la tercera, normativa. Parte del problema de este debate tiene su origen en el hecho de que la pregunta se trata como si fuera empírica, cuando en realidad normalmente es normativa o teórica.

Consideremos por un momento algunas de las cuestiones que plantea la versión empírica de la pregunta. ¿Piensan quienes financian los trabajos que existe una diferencia entre evaluación e investigación? Si así es, ¿de qué modo afecta esta distinción a las decisiones sobre financiación? Las entidades estatales de financiación de investigaciones pueden distinguir, por ejemplo, entre investigación básica,

investigación estratégica e investigación aplicada. No es seguro que «cuenten» como evaluación a cualquiera de ellas. ¿Qué piensan de la distinción los usuarios? No creo, por ejemplo, que a los directores de un gran organismo nacional de voluntarios para la atención de la infancia les importe mucho si algo se considera una evaluación o una investigación, siempre que demuestre tener una «utilidad» clara. Además, ¿qué piensan los profesionales de la evaluación o de la investigación? Si pudiéramos diseñar alguna hipotética doble prueba oculta, ¿quienes la aplicaran podrían distinguir empíricamente los procesos y los productos de la evaluación, por un lado, y los de la investigación, por el otro? ¿Y los encuestados? En este campo, el trabajo empírico puede ser saludable, aunque pueda resolver pocos desacuerdos, porque éstos no se plantean en el nivel de la sustancia, sino en el de los axiomas o de los objetivos. Los desacuerdos axiológicos son intransigentes (Witkin y Gottschalk, 1988).

TABLA 1.1. *Evaluación e investigación.*

Evaluación	Investigación
Trata de problemas prácticos	Trata de problemas teóricos
Culmina en la acción	Culmina en la descripción
Hace juicios de mérito/valor	Describe
Trata temas a corto plazo	Trata temas a largo plazo
Métodos de evaluación	Métodos de investigación
No es disciplinaria	Es disciplinaria
Incluye la evaluación desde dentro	Siempre se realiza desde fuera

La cuestión se complica también por el hecho de que existe poco consenso sobre cómo definir la investigación o la evaluación. Esto a veces lleva a que los protagonistas predefinan ambas de una forma que anuncia la solución.

¿Qué diremos de la tabla 1.1? Sin duda es posible encontrar defensores respetados de cada una de las posturas. Por ejemplo, Ernest House, uno de los comentaristas y teóricos más lúcidos en el campo de la evaluación, acepta una versión de la distinción entre a corto y a largo plazo, porque piensa que la ciencia busca unas generalizaciones a largo plazo para públicos universales, mientras que la evaluación busca unas conclusiones vinculadas al contexto y para un público determinado (House, 1980). Procura distanciarse de Donald Campbell, quien parece defender que la evaluación es parte de la indaga-

ción científica y está sujeta a similares intereses epistemológicos. La distinción que traza House es importante, pero encuentro en ella poco que pueda servir de base verosímil para distinguir entre evaluación e investigación. En el capítulo 4 veremos que los estudios realizados sobre los usos de la evaluación y de la investigación política demuestran que a muchos funcionarios y políticos les influyen las pruebas referentes a más largo plazo, más que los estudios particulares a corto plazo.

La distinción entre teórico y práctico tampoco es convincente. Sin embargo, ha tenido defensores influyentes. Algunos líderes del movimiento de la evaluación democrática en el ámbito de la educación y del movimiento del profesor como investigador aceptan cierta forma de esta distinción. Elliott recuerda el surgimiento de la evaluación democrática en la educación inglesa en los años sesenta y setenta. Parte de los planes de este movimiento era excluir los métodos empiricistas del campo de la evaluación. Para algunos, con esto se llegó incluso a poner en entredicho la continuidad entre cualquier conjunto de métodos de investigación y la tarea evaluativa, y a cuestionar «la suposición de que la pericia evaluadora se pueda definir básicamente desde la perspectiva del dominio de un conjunto de métodos» (Elliott, 1991, pág. 218). El contraargumento de Elliott y de otros es que el *público* es la clave que distingue la investigación naturalista de la evaluación naturalista. En esta última, el público lo constituyen individuos y grupos que tienen un interés práctico en el estudio, en oposición a individuos iguales del mundo académico.

La distinción entre práctico y teórico plantea, sin ninguna duda, cuestiones materiales. Se ha dado peso acertadamente al argumento según el cual la evaluación no debe servir a intereses académicos, sino a los del grupo de referencia que la evaluación considera. Nos fijaremos con detenimiento en este tema a lo largo del libro. Mi problema es doble. En primer lugar, parece innecesario aceptar el supuesto de que las ideas, los conceptos y las teorías son enemigos de la acción comprometida. Los etnógrafos de tendencias generales y los evaluadores abiertamente ideológicos son los dos únicos grupos que palidecerían tan sólo con pensarlo. En segundo lugar, parece que los evaluadores democráticos aceptan la idea asociada de que el evaluador tiene una postura política, mientras que «al investigador no se le exige tal postura. Se sitúa fuera del proceso político, y valora su alejamiento de él» (Elliott, 1991, pág. 219). Una vez más, es una teoría muy simplista, y Elliott ha cambiado de ideas para resolver este dualismo.

Lincoln y Guba ofrecen quizá los argumentos más sólidos a favor de una distinción entre evaluación e investigación. En un artículo retrospectivo, Lincoln recuerda que concluyeron que existen formas diferentes de «indagación disciplinada». Adoptaron esta útil expresión de Cronbach, para referirse a una investigación que tiene una textura que muestra la materia prima que se emplea en la argumentación y en los procesos lógicos y con la que éstos se comprimen y se rediseñan para hacer creíble la conclusión. Concluyeron los autores que la investigación, la evaluación y el análisis de políticas son formas diferentes de indagación disciplinaria. De ahí que «no tenía sentido hablar de “investigación de evaluación”, salvo como investigación *sobre* métodos o modelos de evaluación» (Lincoln, 1990, pág. 76).

Lincoln y Guba definen la investigación como «un tipo de indagación disciplinada con la que se pretende resolver algún problema para lograr una comprensión o facilitar la acción» (1986b, pág. 549). Al construir su definición de evaluación proponen algunas distinciones. En primer lugar, aceptan una distinción familiar de Scriven entre evaluación sumativa y formativa. Utilizan el término norteamericano *evaluand* («evaluando») para referirse a la entidad que se evalúa:

El objetivo de la evaluación formativa es proporcionar una información descriptiva y de juicios con la que se perfeccione, se mejore, se altere y/o se modifique el evaluando, mientras que el objetivo de la evaluación sumativa es determinar sus efectos, sus resultados o sus consecuencias (1986b, pág. 550).

En segundo lugar, distinguen dos aspectos que el evaluador puede querer establecer, el *mérito* y el *valor*. El «mérito», tal como ellos emplean el término, se refiere a las cualidades intrínsecas y ajenas al contexto que acompañan al evaluando de un lugar a otro y que son relativamente invariables. Por ejemplo, un currículo puede tener un valor independiente de cualquier aplicación local que de él se haga. El «valor» es aquel que determina el contexto y que varía de uno a otro. Por ejemplo, un currículo puede tener cierto valor para enseñar a un determinado niño en un enclave particular. Basándose en estas distinciones, definen la evaluación como

un tipo de indagación disciplinada que se realiza para determinar la virtud (el mérito y/o el valor) de una determinada entidad, el evaluando —por ejemplo, un tratamiento, un programa, unos servicios, una actuación, etc.—, con el fin de mejorar o perfeccionar al evaluando

(evaluación formativa) y valorar su impacto (evaluación sumativa) (1986b, pág. 550).

Estas definiciones distintivas tienen mucho a su favor. Sorteando de forma segura la tendencia al estereotipo y las argucias que caracterizan a muchos esfuerzos por establecer esas distinciones. Evitan hacer de la metodología el criterio para la distinción y, además, hacen de la diferencia de propósito el tema fundamental y más importante: «La investigación se realiza para resolver algún problema, mientras que con la evaluación se pretende crear valor» (1986b, pág. 551).

Hay, sin embargo, unos ecos persistentes de tinte retórico que de nada sirven. Pese al cuidado comparativo con que se elaboran las definiciones, los comentarios que las acompañan vuelven a incluir algunas formas de distinguir la investigación de la evaluación que no se deberían aceptar. No tiene sentido decir, por ejemplo, que «el resultado o el uso esperados de la investigación es simplemente la expansión del conocimiento» (1986b, pág. 551) mediante lo que Lincoln y Guba llaman «informes técnicos». Ni basta con que digan que es más probable que el público de la evaluación sea múltiple, heterogéneo e incluya a grupos de interés, mientras que el público de la investigación es relativamente homogéneo. Además, decir que los informes de evaluación «hacen unas consideraciones políticas en una medida que sencillamente no se encuentra en la investigación» (1986b, pág. 552) es una flagrante y exagerada simplificación. Las ventajas de haber evitado las distinciones metodológicas en sus definiciones básicas se pierden cuando después parece que dicen que las correspondencias entre investigación, evaluación y análisis de política sólo se producen cuando las técnicas que se emplean para llegar a los datos, y para sacar conclusiones de ellos, sólo «se solapan ocasionalmente» (1986b, pág. 560). Más en general, un punto débil recurrente en todos los escritos de Lincoln y Guba es que dan la falsa impresión de que la mayoría de los investigadores convencionales siguen lo que ellos denominan el «paradigma tradicional» (1986b, pág. 555).

En resumen, existen diferencias importantes entre la evaluación y la investigación que derivan de los distintos grupos de propósitos que dirigen cada forma de indagación. Pero las generalizaciones exageradamente simplistas sobre las entidades homogéneas de «investigación» y de «evaluación» son la base de la retórica, y no de la práctica. En mi opinión, hablar de «investigación de evaluación» o de «investigación de políticas» tiene completo sentido, y no supone confusión alguna de categorías.

LA EVALUACIÓN CUALITATIVA

¿Qué afirmamos cuando decimos que una evaluación es cualitativa? ¿Cuáles son las características de identidad de un enfoque cualitativo? ¿Cuáles son sus virtudes frente a los métodos cuantitativos? Antes de comprometernos con algunas respuestas, debo decir que este libro no pretende simplemente describir los métodos cualitativos de evaluación, sino también, y en primer lugar, establecer y analizar la naturaleza de la metodología de la evaluación cualitativa. Empleo el término metodología en un sentido muy amplio en este contexto, para incluir la interacción entre la práctica de los métodos cualitativos, las bases epistemológicas de la evaluación cualitativa, y el propósito evaluativo. Por consiguiente, «lo que distingue de forma importante a una metodología de evaluación de otra no son los métodos, sino qué preguntas se abordan y qué valores se fomentan» (Greene, 1994, pág. 533).

A modo de preámbulo, debemos reconocer que existe una relación estrecha, y excesivamente delicada, entre la comprensión cualitativa y el conocimiento de sentido común. Por conocimiento de sentido común no me refiero a lo que Geertz denomina «una mera aprehensión de hecho de la realidad» (1983, pág. 75). El influyente artículo de Geertz convence al lector con varias razones de por qué el hecho de tratar el sentido común como un «cuerpo relativamente organizado de pensamiento considerado» —«en resumen, un sistema cultural»— y no simplemente como «lo que cualquiera en su sano juicio sabe» debe desembocar en algunas conclusiones útiles (1983, págs. 75, 76).

Si se acepta esta visión de las cosas cotidianas, aparecen unas consecuencias inmediatas y extensivas para la forma de entender la evaluación cualitativa. Hemos dedicado cierto tiempo a considerar la relación que existe entre investigación y evaluación, y hemos señalado que Lincoln y Guba sitúan la cuestión en un contexto más amplio al hablar de diversas formas de indagación disciplinada. Pero si se incluye el sentido común, el contexto es mucho mayor. Campbell, por ejemplo, insiste en que «no debemos suponer que el conocimiento científico sustituye al conocimiento de sentido común. Al contrario, la ciencia depende del sentido común aunque, en el mejor de los casos, lo trascienda» (1978, pág. 186). Indica el autor que el conocimiento corriente y la mayor parte de las formas de ciencia coinciden en algún punto entre los extremos del completo escepticismo (en el que abandonamos el «conocer» o la ciencia) y la credulidad absoluta.

«De algún modo combinan una capacidad para la desconfianza y la revisión sobre determinados hechos con una creencia en un cuerpo común de afirmaciones de conocimientos» (1978, pág. 187).

Eisner (1991a, pág. 15) traza una útil distinción entre la investigación cualitativa, la indagación cualitativa y el razonamiento cualitativo. Es una distinción útil porque implícitamente plantea la cuestión de cómo deberíamos entender la relación entre investigación, evaluación y propósito y práctica profesionales, y de qué forma la importancia de las ideas sobre el conocimiento cualitativo se extiende a la vida cotidiana.

Las características de la indagación cualitativa

Cualquier intento de hacer una relación de aquello que es común en la evaluación cualitativa no contará con un acuerdo universal, ni siquiera mayoritario. En efecto, existe la idea muy extendida de que cualquier pretensión de caracterizar la investigación cualitativa en su conjunto está abierta a severas críticas (Silverman, 1997b). Confío evitar este debate. No pretendo ofrecer una lista de lo que sea genérico de toda investigación cualitativa, ni siquiera exhaustivo de la evaluación cualitativa. Al contrario, digo que estas características son esencialmente relevantes en la evaluación cualitativa. No todo el mundo las va a aceptar. Por ejemplo, veremos en el capítulo 3 que en las cuestiones aparentemente más prosaicas sobre los aspectos específicos de la investigación sobre la evaluación resuenan disputas sobre si la evaluación cualitativa tiene una base realista o relativista. La lista que aquí se presenta tiene un trasfondo realista. Toma como punto de partida la obra de Miles y Huberman.

- La evaluación cualitativa se realiza a través de un contacto intenso y a largo plazo con el campo: «Estas situaciones son típicamente “banales” o normales, reflejo de la vida cotidiana de los individuos, los grupos, las sociedades y las organizaciones» (Miles y Huberman, 1994, pág. 6). Implica «fijarse en lo corriente en los lugares donde adquiere unas formas desacostumbradas, de manera que comprender la cultura de unas personas pone al descubierto su normalidad sin reducir su particularidad» (Geertz, 1973, pág. 14).
- La función del investigador es alcanzar una perspectiva general «holística» de la cultura y el contexto objeto de estudio.

- El holismo se consigue mediante la indagación en lo *particular*. «Lo característico es que el antropólogo se plantee [...] unas interpretaciones más amplias [...] a partir de una familiaridad excesivamente ampliada con asuntos extremadamente pequeños.» Las grandes realidades del Poder, la Fe, el Prestigio, el Amor, etc., se afrontan «en unos contextos lo suficientemente oscuros [...] para eliminar las mayúsculas» (Geertz, 1973, pág. 21). Las evaluaciones cualitativas «hacen palpables los argumentos» (Eisner, 1991a, pág. 39).
- El todo y lo particular están en tensión. «Los hechos pequeños hablan de grandes temas» (Geertz, 1973, pág. 23), y «en lo particular se sitúa un tema general» (Eisner, 1991a, pág. 39). Sin embargo, a veces nos queda la impresión de que este proceso es completamente obvio, de que lo que comprendemos sobre unas aulas determinadas, unos determinados trabajadores sociales, unas determinadas clases prácticas de medicina, etc., inmediatamente será útil para comprender otras aulas, otros trabajadores sociales y otras clases prácticas. Sin embargo, el trabajo de los evaluadores cualitativos ha sido deficiente en este sentido. Greene se queja de que

las aportaciones potenciales de la historia a una comprensión científica social más general quedan acalladas por el rechazo de los evaluadores cualitativos de las explicaciones formales y proposicionales de los fenómenos sociales. Al mismo tiempo, también se le ha puesto sordina a la voz de los evaluadores cualitativos en el foro de la política, y las historias que contamos sólo tienen una fuerza local (1996, pág. 281).

- «El investigador pretende conseguir datos sobre las percepciones de los actores locales “desde dentro”, a través de un proceso de profunda cortesía, comprensión empática (*verstehen*), y suspensión o “puesta entre paréntesis” de las percepciones sobre los temas de que se habla» (Miles y Huberman, 1994, pág. 6). Hay que hacer una advertencia. Esta postura no excluye una posición normativa. En efecto, los enfoques cualitativos «pueden dar voz efectiva a lo que normalmente está silenciado, y pueden iluminar de forma conmovedora lo que habitualmente está oculto» (Greene, 1994, pág. 541).
- Las categorías del encuestado o del miembro se mantienen en primer plano a lo largo de la evaluación.

- La evaluación cualitativa es interpretativa: «Una tarea importante es explicar la forma en que las personas de determinados enclaves llegan a comprender, justificar, gestionar de cualquier otro modo sus situaciones cotidianas, y actuar» (Miles y Huberman, 1994, pág. 7). Por consiguiente, «los datos cualitativos no son tanto sobre la “conducta” como sobre las *acciones* (que conllevan intenciones y significados, y producen unas consecuencias)» (1994, pág. 10). Esto no significa que la conducta no sea importante. «Hay que atender también a la conducta, y con cierta precisión, porque es a través del flujo de la conducta —o, más exactamente, de la acción social— como encuentran articulación las formas culturales» (Geertz, 1973, pág. 17).
- Se utilizan relativamente pocos instrumentos estandarizados, sobre todo al principio. El investigador es el principal instrumento del estudio. El trabajo de campo cualitativo no es sencillo: «Las características que cuentan en un determinado enclave no llevan la etiqueta puesta» (Eisner, 1991a, pág. 33).
- Por último, «gran parte del análisis se hace con palabras» (Miles y Huberman, 1994, pág. 7). Intervienen mucho el juicio y la persuasión mediante la razón, y en la evaluación cualitativa los hechos nunca hablan por sí mismos. Los evaluadores cualitativos suelen decir que una buena evaluación se mide, al menos en parte, por su coherencia. Geertz disiente por completo de esta idea y propugna una visión más realista de las explicaciones cualitativas. También nosotros disentimos, sin por ello querer negar la omnipresencia de la estética en las exposiciones cualitativas de las evaluaciones.

Nada hay más coherente que la falsa impresión de un paranoico o la historia de un timador. La fuerza de nuestras interpretaciones no puede radicar, como tan a menudo se pretende hoy, en lo ajustado de su cohesión (Geertz, 1973, pág. 18).

Para defender las virtudes que la indagación cualitativa aporta a la evaluación, es mejor poner ejemplos que afirmarlas. Pero la atención a los sucesos que se producen de forma natural, la base local y la contextualización de la evaluación, la flexibilidad del diseño del estudio, cualquier posibilidad de desvelar la complejidad mediante enfoques holísticos tienen un valor en cierto grado evidente. Además, los datos cualitativos tienen algunas virtudes que los evaluadores no siempre aprovechan. Por ejemplo, será un tema recurrente en este li-

bro que las evaluaciones cualitativas pueden valorar la causalidad «tal como actúa realmente en determinados enclaves» (Miles y Huberman, 1994, pág. 10).

Williams (1986) recopila los criterios por los que se puede decidir si una evaluación cualitativa o naturalista es apropiada. Será apropiada en circunstancias en que:

- Los temas de la evaluación no estén claros de antemano.
- Las definiciones oficiales del evaluando no sean suficientes, y se necesiten unas perspectivas desde dentro («émicas»).
- Se requiera una «descripción gruesa» (expresión de Geertz).
- Sea deseable que se transmita el potencial de experiencia indirecta del evaluando por parte del lector.
- Sea apropiada la evaluación formativa, destinada a mejorar el programa, la política o la práctica.
- El resultado incluya complejas acciones en enclaves naturales.
- Los destinatarios de la evaluación quieran saber cómo actúa el evaluando en su estado natural.
- Haya tiempo para estudiar al evaluando a través de su ciclo natural. La auténtica fuerza de la evaluación naturalista se desperdicia si no se dispone de tiempo para observar las funciones naturales del evaluando en sus diversas formas.
- La situación permita una indagación intensiva, sin plantear graves obstáculos éticos.
- Se pueda estudiar al evaluando discretamente, tal como actúa, y de forma ética.
- Se disponga de diversas fuentes de datos.
- Se cuente con los recursos y el consentimiento para reunir ejemplos y pruebas negativas.
- Exista un suficiente acuerdo con el cliente y el usuario sobre la estrategia metodológica.

Se me disculpará que piense que son éstas toda una batería de razones en favor de la evaluación cualitativa. En efecto, la idea recurrente de este libro es que la evaluación cualitativa normalmente será la metodología elegida. Es una razón más de por qué el planteamiento pluralista del «caballo según carrera» no es adecuado. En un influyente estudio de los primeros tiempos se examinaba un caso en que el análisis de los datos cualitativos daba una explicación que no se podía conciliar con los datos cuantitativos del mismo estudio. Trend reflexiona sobre el caso con la esperanza de que «pueda ayu-

dar a disipar la idea de que con el uso de métodos múltiples se conseguirán explicaciones más sólidas de forma fácil y aditiva» (1979, pág. 68). Señala el autor la idea común de que los datos observacionales se han de emplear para generar hipótesis o describir procesos, mientras que los datos cuantitativos se deben utilizar para analizar resultados y verificar hipótesis. Concluye Trend:

Creo que esta división del trabajo es rígida y restrictiva [...] No siempre es evidente la complementariedad. Limitarse a utilizar perspectivas diferentes, con la esperanza de que se validen mutuamente, no explica qué habrá que hacer si las piezas no encajan (1979, pág. 83).

Los argumentos a favor de la evaluación cualitativa frente a la cuantitativa no se deben sacar de su contexto. En el capítulo 3, en mi crítica a lo que se dice sobre el paradigma, quedará claro que rechazo la idea de que la indagación cuantitativa y la cualitativa no guardan relación. En este sentido, coincido con Silverman cuando dice que «no existe una base de principios para adoptar un enfoque o cualitativo o cuantitativo. Todo depende de lo que se pretenda hacer» (1997b, pág. 14). Mi tesis es que existen las mismas probabilidades de que lo que pretendemos hacer —el propósito de la evaluación— exija o no un enfoque cualitativo, un hecho que se ve reforzado por la necesidad de una redistribución correctora a la luz de las ansias por conseguir resultados y pruebas, y poder dar razón de lo que se haga.

SINOPSIS

Iniciaba el apartado anterior diciendo que por metodología entiendo la interacción entre la práctica de los métodos cualitativos, las bases epistemológicas de la evaluación cualitativa, y el propósito evaluativo. Al considerar las creencias sobre la naturaleza del conocimiento y de la ciencia que informan la indagación cualitativa, consideraré (capítulo 3) tres posturas: el realismo pospositivista, el constructivismo relativista y la evaluación crítica. La triple distinción es un tanto idealizada. Por ejemplo, en los tres casos se acepta de diferente forma que los datos están contruidos socialmente, pero los constructivistas ponen este argumento en primer plano. Los evaluadores críticos y muchos constructivistas son también realistas, pero dudan (bien que por razones muy distintas) de la posibilidad de alcanzar una aproximación ajustada a la realidad mediante la evaluación que

llevan a cabo. En las explicaciones constructivistas están presentes cuestiones estructurales, pero los defensores de la evaluación crítica las ponen en primer plano, además de estar firmemente convencidos del carácter moral de la tarea evaluativa y de adoptar a menudo una postura fuertemente partisana sobre el uso de los conocimientos. Como ya se verá, no me gusta la retórica de los paradigmas, aunque es posible que, siempre que en la evaluación aparezca un ingenuo pragmatismo metodológico, sienta la misma consternación, o quizá mayor.

A lo largo del capítulo 4 nos fijaremos en cuestiones más «orientadas al uso»:

- ¿Cómo sabemos si hacemos bien la evaluación?
- ¿Los resultados de la investigación de evaluación se utilizan mediante la aplicación directa a los problemas, o mediante un proceso de explicación más general?
- ¿Qué principios éticos y qué normas de decisión guían la evaluación cualitativa?
- ¿Cómo deben configurar las respuestas a estas preguntas la metodología que empleemos?

Son preguntas no menos espinosas que las que se consideran en el capítulo 3. En cada caso, es su propósito evaluativo el que las configura, las limita o las enriquece. Las cuestiones de la validez de la indagación, los usos que se le pueden dar a la indagación evaluativa, el planteamiento general de los métodos de investigación, e incluso, en cierta medida, las cuestiones de la ética de la evaluación, adquieren un rasgo distintivo por su propósito evaluativo. Con ello no se pretende ningún elitismo territorial. No digo que los temas de la evaluación sean más complejos, más importantes ni más interesantes que las correspondientes cuestiones de la investigación cualitativa. Pero son *diferentes*.

Después de examinar las inquietudes de lo que Becker llama la preocupación filosófica y metodológica como profesión, en los capítulos 5 y 6 cambiamos de enfoque para analizar la idea de que la materialización del propósito evaluativo no se debe considerar exactamente la misma cuando se trata de políticas, de programas o de profesionales individuales. En el capítulo 5 reflexionaremos sobre algunos de los temas que emergen cuando nos embarcamos en estrategias cualitativas de evaluación de programas. En la segunda parte de ese capítulo explicamos en detalle las ideas alternativas sobre cómo

se puede utilizar la investigación de políticas. En particular, analizamos el debate sobre los modelos instrumentales *frente* a los explicativos del uso de la información. En el capítulo 6 se abordan algunas de las interesantes cuestiones que plantea el reciente y vigoroso incremento del interés y la práctica en el campo de la evaluación directa de la práctica. Consideraremos cuatro áreas de práctica de la evaluación que se pueden resumir bajo el amplio título de evaluación del profesional.

1. Investigación y evaluación realizadas por los profesionales.
2. Investigación participativa.
3. Evaluación como una dimensión de la práctica directa, a través, por ejemplo, de la indagación reflexiva.
4. La evaluación *para* profesionales normalmente realizada por investigadores académicos.

Después de analizar las especiales consideraciones que surgen al realizar una evaluación cualitativa de programas, de políticas y de la práctica individual, podemos concluir con toda seguridad que las diferencias entre ellas están lejos de ser superficiales o marginales. Pero estas diferencias no justifican la fragmentación de la investigación de evaluación y de políticas en múltiples formas discretas de indagación, cada una con sus propios axiomas y su metodología. Desde el capítulo 7 hasta el final del libro pretendo ilustrar cómo los propósitos evaluativos dan una cohesión a unos enfoques diferenciados pero coherentes de las decisiones sobre el diseño, el trabajo de campo y el análisis de la evaluación cualitativa. Las decisiones de diseño cualitativo implican:

- La valoración de la evaluabilidad del evaluando.
- Cuestiones de ética de diseño y de justicia social.
- Diseñar una evaluación cualitativa con el fin de comprender los resultados.
- La utilización de los diseños cualitativos de estudio de casos.
- Diseñar una evaluación participativa.

Mi objetivo en el capítulo 8 es ofrecer una introducción ilustrada a una perspectiva evaluativa sobre el trabajo de campo, más que un inventario resumido de métodos de trabajo de campo que en otras partes de la literatura se abordan con mayor profundidad. En la parte principal de ese capítulo consideraremos de qué forma los propó-

sitos evaluativos configuran las decisiones de utilizar entrevistas, métodos de observación, historias vitales, documentos, entrevistas de grupo y simulaciones. Cualesquiera que sean los métodos, hay unas cuestiones metodológicas sobre el trabajo de campo recurrentes, de las que nos ocupamos en apartados posteriores de ese capítulo. Entre ellas están los temas relacionados con el acceso, los grupos de evaluación sensibles, la evaluación de la satisfacción del interesado y del usuario del servicio, la evaluación del profesional y la evaluación participativa para conferir mayores competencias. Concluimos ese capítulo recordando que las cuestiones éticas no sólo se plantean al inicio de la evaluación cualitativa, sino a lo largo de todo el trabajo de campo.

La mayor parte del último capítulo trata de los temas y las cuestiones que plantea el análisis de la evaluación cualitativa. Retomamos uno de los hilos del capítulo 6 y señalamos de qué forma la evaluación participativa y del profesional configura el enfoque del análisis. El análisis participativo plantea importantes cuestiones de ética del análisis, y nos ocupamos de éstas y otras consideraciones que surgen en las fases de análisis y de informe de la evaluación. Así se delimita un contexto para considerar el tema esencial del capítulo: cómo fundamentar y apoyar lo que en la evaluación se diga. Recogemos la argumentación del capítulo 7 referente a la relevancia de la evaluación cualitativa para abordar las espinosas cuestiones de las causas. De ahí pasamos a considerar algunas afirmaciones sobre cómo se puede hacer más verosímil el análisis cualitativo. Entre ellas están el uso de métodos de validación de los miembros, la triangulación y métodos de síntesis y de metaanálisis. Por muy importantes y difíciles que puedan parecer estas cuestiones, probablemente lo sean más los temas relacionados con la generalización y la validez externa de la evaluación. Señalamos y revisamos las opciones y los enfoques más ampliamente estudiados de estos dos temas. Concluimos con una última visita a algunos de los temas sobre la autoría, el informe y el uso de la evaluación de los que nos ocupamos desde una perspectiva más general en el capítulo 4. Pero antes de adentrarnos en la tesis principal del libro, es obligado un preámbulo. Las razones de la evaluación cualitativa surgen en el contexto del trabajo animado y continuado de los teóricos de la evaluación. En el capítulo 2 se esbozan las líneas principales de este trabajo teorizador.

CAPÍTULO 2

TEÓRICOS DE LA EVALUACIÓN Y EVALUACIÓN CUALITATIVA: DE PERSUASIONES Y PERSUASIVOS

Este capítulo trata de los «grandes nombres» de la teoría de la evaluación. Hacer una relación de quiénes deben figurar en la lista y quiénes no, es una tarea arriesgada. Lo que para mí es fundamental, para otro puede ser secundario. La elección resulta un poco más fácil por el hecho de que mi principal objetivo es dilucidar cómo se ha producido el desarrollo de la evaluación cualitativa y cómo se ha configurado a través del trabajo de los teóricos de la evaluación. Mi objetivo secundario es esbozar un mapa de los temas y los avances de la teoría de la evaluación más general. Conviene hacer dos sanas advertencias. Primera, esto no va a dar una imagen «prospectiva» de la historia de la evaluación. Los teóricos más recientes albergan serias dudas sobre la imagen de la ciencia como un avance acumulativo y progresivo. Compartimos esas dudas. Los desarrollos particulares pueden enriquecer o empobrecer, ampliar o reducir. Segunda, son pocos los comentarios detallados sobre la narrativa o el relato. No evalúo a los evaluadores. Los temas que plantean se dejan «en suspenso», para recogerlos después a lo largo del libro.

La mayoría de las personas cuya obra se expone en este capítulo no se describirían a sí mismas como teóricos de la evaluación cualitativa, y quizá la mitad de ellas preferirían que no se les tuviera simplemente como teóricos de la *evaluación*. Sin duda podría haber escogido a otras personas cuya influencia sobre la práctica cotidiana de la evaluación fuera al menos tan fuerte como la de quienes aquí figuran. Pero, a mi juicio, estas personas y las teorías que representan son las más relevantes para nuestros propósitos. Campbell y Scriven han configurado la teoría de la evaluación durante cuatro décadas. La obra de Campbell en particular es mucho más esencial para lo que nos interesa de lo que podrían esperar quienes le conocen sólo como el sumo sacerdote del diseño de la investigación experimental y cuasi-experimental. Cronbach, aunque menos conocido, puede pre-

sumir mejor que nadie de defender una *teoría* de la evaluación, mientras que Stake fue el primer gran teórico de la evaluación que se comprometió con una metodología cualitativa. Weiss ha tenido menos que decir sobre la metodología de evaluación, pero mucho sobre el contexto político en que tienen lugar la investigación y la evaluación de políticas.

House da un paso significativo más allá de la política pluralista de sus predecesores, y defiende una idea reformista de la evaluación apoyada en un compromiso con la justicia social. La evaluación no cuenta aún con una teoría desarrollada de la evaluación crítica, abiertamente comprometida con una agenda emancipadora. Pero la crítica está bien articulada y yo señalo sus términos.

Los profesionales que trabajan en el ámbito de la asistencia social, en la educación o en otras agencias locales son más proclives a situar sus cuestiones sobre la evaluación muy cerca de su práctica directa. Reid (asistencia social) y Eisner (educación) representan unos planteamientos de la evaluación muy diferentes de los intereses de evaluación de programas mostrados por la mayoría de las personas de las que previamente se habla en este capítulo. A ambos les interesa la evaluación del profesional, aunque avanzan en direcciones distintas. El capítulo concluye recogiendo la obra de Stake, de tendencia interpretativa y naturalista, y mostrando la forma constructivista radical que adquiere en la obra de Lincoln y Guba.

De esta lista de nombres se puede concluir que la evaluación se inventó y se desarrolló en Estados Unidos. Así es, sin duda, en el caso de la evaluación de programas. La evaluación del profesional, la evaluación democrática y la evaluación participativa han surgido de avances paralelos, y a veces relacionados, producidos en Estados Unidos, Gran Bretaña y, más recientemente, en Australia y Canadá. La evaluación crítica puede encajar en un contexto intelectual inglés al menos tanto como en un contexto norteamericano. En cualquier caso, es posible que deba mucho a raíces intelectuales francesas y alemanas.

¿Las obras de Campbell, de Scriven y demás representan modelos, o tal vez paradigmas? Nuestra opinión personal es reflejo de la de Stake. Sobre los modelos, dice éste que el término «promete en exceso. Las personas empiezan por confiar en él como en un plano, con el “inventario de piezas” y unas “instrucciones de montaje”, de manera que, si se sigue con cuidado, se puede ensamblar una evaluación satisfactoria». En cambio, «nuestros llamados modelos son simplemente unas propuestas que aluden a unas pocas de las muchas responsa-

bilidades, e identifican unas pocas de las muchas oportunidades de indagación» (Stake, 1991, pág. 71). De las pretensiones paradigmáticas podríamos coincidir con la observación de Reid: «otro ejemplo más en el que el nombre de Thomas Kuhn fue tomado en vano» (1990, pág. 141). Quizás, empleando el término de House, se describan mejor como «persuaciones», y en la mayoría de los casos, unas persuaciones sólidas.

DONALD CAMPBELL: INVESTIGAR LAS CAUSAS

Campbell ejemplifica dos compromisos gemelos con la comprensión de las posturas de aquellos con los que puede disentir y con el fomento activo de la crítica de su propia postura. Tal vez se le conozca más por su trabajo metodológico sobre el diseño de la investigación experimental y cuasi-experimental, y por su formulación de las amenazas a la validez de lo que él describe como estudios de «investigación de causas» (Campbell y Stanley, 1966; Cook y Campbell, 1979).

En los años sesenta, debido en parte a su influencia, se presumía ampliamente que los diseños experimentales para medir los resultados de los programas innovadores eran el punto de referencia y la prueba fundamental de *toda* evaluación. Campbell acuñó la expresión «validez interna» para referirse a las interferencias sobre la relación entre variables independientes y dependientes. Da un mayor peso a la validez interna que a la externa. ¿Qué utilidad tiene, diría, generalizar una relación si se duda de la propia relación? Su compromiso político más amplio ha sido con un mundo emancipador conseguido mediante una sociedad experimentadora.

Campbell es un realista crítico que acepta la existencia de un mundo real más allá de los conocedores, pero que critica con energía el menor signo de realismo ingenuo que presuma que nuestras observaciones o nuestras teorías reflejan directamente la realidad. El énfasis que pone en la imposibilidad lógica de la inducción y la primacía de los métodos de falsación para justificar las pretensiones de conocimiento, unido a su realismo crítico, revelan en su conjunto su deuda con Karl Popper.

Partió de sus primeros trabajos sobre las diferencias culturales en el grado de susceptibilidad ante las ilusiones ópticas, para explorar de qué forma los fracasos de comunicación transculturales se podrían distinguir de las diferencias perceptuales transculturales. Su con-

clusión es que «es extremadamente difícil afirmar que los miembros de otra cultura perciban las cosas de manera distinta» (Campbell, 1964, pág. 325). Al reflexionar sobre estos estudios más de treinta años después, insiste en que «interpretar como una diferencia cultural lo que en realidad es un fracaso de comunicación [es] la más omnipresente fuente de error en los esfuerzos por conocer al otro» (Campbell, 1996, pág. 165). Sin embargo, no cree que estas dificultades sean insuperables. «Quienes hacen difícil el conocimiento del otro están en lo cierto. Quienes lo consideran imposible en cualquier grado están equivocados» (1996, pág. 169), de modo que rechaza a los relativistas («nihilistas ontológicos» es su expresión) y «otras importaciones más recientes desde los cafés de París» (1996, pág. 159). Mantiene, pues, una epistemología de doble filo, insistiendo en la inevitabilidad de la duda, pero también en que podemos superarla hasta un nivel prácticamente útil y conocer al otro.

El etnógrafo Howard Becker impartió clases con Campbell entre los años 1960 y 1979 e influyó mucho en dos ensayos que éste escribió en los años setenta (Campbell, 1978, 1979). Campbell llegó a aceptar las críticas que Becker hacía de la ciencia social cuantitativa estándar como «válidas y extremadamente relevantes» (Campbell, 1996, pág. 161) y criticaba a los científicos cuantitativos que, «bajo la influencia de misioneros del positivismo lógico, presumen que, en la auténtica ciencia, el conocimiento cuantitativo sustituye al conocimiento cualitativo y de sentido común» (Campbell, 1979, pág. 50). Considera que el conocimiento cualitativo es tanto la prueba como el componente básico del conocimiento cuantitativo. «Esto no significa decir que tal observación naturalista de sentido común sea objetiva, fiable e imparcial. Pero es todo lo que tenemos. Es el único camino al conocimiento, por ruidoso, falible y tendencioso que pueda resultar» (1979, pág. 54).

Su segunda aportación a la evaluación cualitativa fue suscitar la cuestión de si los métodos cualitativos pueden cumplir el mismo objetivo de explorar las causas que los métodos cuantitativos. Volveré a sus argumentaciones en el capítulo 7, pero baste decir que Campbell emplea la analogía de la investigación experimental para sugerir formas de fortalecer los estudios de casos y «construye uno de los argumentos más sólidos hasta la fecha sobre la capacidad de los métodos cualitativos para producir conocimientos causales válidos» (Shadish y otros, 1990, pág. 135). Más recientemente ha señalado que, «además del enfoque del estudio de casos cuantitativo y cuasi-experimental [...] en nuestro armamento de las ciencias sociales se necesita

también un método de estudio de casos humanístico y que busque la validez» (Campbell, 1994, pág. x).

La obra de Campbell se ha apreciado de nuevo en los últimos diez años, y sigue influyendo de forma significativa en la investigación cualitativa. Se han hecho críticas importantes a la aparente ambigüedad de sus ideas sobre la causalidad, su abandono de los métodos de investigación de «descubrimiento», la prioridad que otorga a la validez interna y el carácter utópico de su visión social. Observaremos el desarrollo de estas críticas en la obra de Cronbach, Stake, House y Weiss. Sin embargo, Campbell ofrece una agenda continuada para los profesionales de la evaluación cualitativa. Una agenda que incluye su

- elaborada defensa del realismo crítico,
- trabajo relativo a los estudios de casos y las etnografías comparativas,
- aceptación del comentario crítico.

MICHAEL SCRIVEN: UNA CIENCIA DE LA EVALUACIÓN

Al igual que Campbell, Scriven pone el énfasis en la evaluación como búsqueda de soluciones eficaces a los problemas sociales. Sus ideas han sido objeto de debate en casi todos sus puntos, y el autor ha desarrollado «uno de los enfoques de los que más se habla y que menos se usan» (House, 1980, pág. 232). Ha introducido un lenguaje nuevo en el campo —la evaluación sin metas, la evaluación formativa y sumativa, la metaevaluación y la inferencia probatoria—. La evaluación, incluida la evaluación cualitativa, ha estado configurada por la obra de Scriven, incluso cuando ésta se aleja de manera más evidente de aquélla. Las ideas de Scriven tienen una importancia generativa en el campo, «y quizá no se aprecien aún en todo su valor» (Shadish y otros, 1990, pág. 74).

Le interesa menos desarrollar y prescribir una metodología de la evaluación que explorar las exigencias lógicas de la tarea. En la base de su tesis está la idea de que la evaluación asigna mérito o valor. Para Scriven, la misma idea de unos evaluadores que nunca sacan unas conclusiones evaluativas «no sólo es paradójica, sino que hace superflua la profesión autónoma de los evaluadores. No son más que científicos sociales con otro nombre» (Scriven, 1995, págs. 54-55). «El hecho de que la evaluación, tal como la mayoría de sus profesionales la conciben hoy, sólo pueda llevar a conclusiones descriptivas, y no a

conclusiones evaluativas» representa, en su opinión, «la enfermedad terminal» de la evaluación (Scriven, 1986, pág. 10). Sin embargo, distingue de forma cuidadosa entre los juicios sobre el mérito de lo que se evalúa (el «evaluando», para emplear el término comúnmente usado en la literatura norteamericana), las *explicaciones* (mediante la evaluación formativa), y las *recomendaciones* («remediaciones»):

Lo malo es malo y lo bueno es bueno, y corresponde a los evaluadores decidir qué es cada cosa. Y hay muchas ocasiones en que deberían decir qué es cada cosa, dispongan o no de explicaciones o remediaciones (Scriven, 1986, pág. 19).

La «evaluación sin metas» es para Scriven el medio ideal para conseguir un juicio evaluativo, con lo cual se refiere a que los evaluadores deben actuar tras un velo de ignorancia sobre las metas que se propongan los gestores del programa y otros grupos de intereses. Los objetivos predisponen al evaluador. En vez de preguntar: «Éstos son los objetivos, ¿se han conseguido?», Scriven preferiría decir: «Éste es el programa, ¿cuáles son sus efectos?». Al ignorar las metas del programa, el evaluador buscará *todos* los resultados. Scriven rechaza la distinción entre efectos buscados y efectos secundarios, porque refleja las intenciones del productor. Los criterios para juzgar los programas proceden de las valoraciones de las necesidades; por consiguiente, el enfoque de nuestro autor parte del consumidor y no del gestor. Esto conduce a una práctica de la evaluación en la que el principal evaluador es casi inevitablemente alguien externo: «Tanto el distanciamiento como la objetividad siguen siendo unos ideales correctos y a menudo alcanzables para el evaluador externo» (Scriven, 1997, pág. 483). «La distancia tiene su precio», pero la implicación «pone en peligro todo el capital», y «el llamado diseño participativo [...] es tan descuidado como uno sea capaz de hacerlo» (Scriven, 1986, págs. 488, 486).

Scriven pretende integrar gran parte de estas tesis en una lógica general de la evaluación. Cree que, sin una teoría general de este tipo, la evaluación se pierde en la estrechez y la carencia de perspectiva, incapaz de abordar las batallas internas y externas a las que se enfrenta (Scriven, 1996). Su objetivo es una teoría de la evaluación de amplio alcance, que contemple no sólo la evaluación de programas, sino también la evaluación de productos y de personal. Entiende la evaluación como una «transdisciplina», autónoma tanto respecto a la filosofía como respecto a la ciencia social, que debe ofrecer un

proceso de reflexión apropiado para todas las disciplinas en la vigilancia de sus fronteras disciplinares.

En su aspecto más negativo, parece que Scriven va haciéndose cada vez más maniático, tanto en los contenidos como en el tono. Su intolerancia de la evaluación formativa, los constructivistas, los «filosóficamente obsesos» (Scriven, 1997, pág. 479) y lo que él denomina la visión «lacaya» de la investigación de evaluación (Scriven, 1986, pág. 23) parecen como la «bofetada de autoridad» que suelta el «anciano de la tribu», que cuenta con «licencia para la irascibilidad» (Kushner, 1997, pág. 364). Es lamentable. Scriven es un pensador original que repetidamente pone en entredicho lo que otros dan por supuesto. House, por ejemplo, dice que Scriven «ha escrito cosas brillantes sobre la objetividad» (1980, cap. 4; Scriven, 1972) y sus últimos trabajos sobre la aplicación de una lógica informal al razonamiento evaluativo han dado origen a obras creativas sobre diversos frentes (por ejemplo, Fournier, 1995). Por último, «su obra sobre la valoración es tan compleja, sutil y contiene tanta información que cada vez que se lee se aprenden cosas nuevas» (Shadish y otros, 1990, pág. 118). Pero su modelo general de evaluación no es para el mundo real y, en el mejor de los casos, será parcial y selectivo en sus consecuencias para la evaluación cualitativa.

LEE CRONBACH: EVALUAR DENTRO DE LOS PROGRAMAS

Los temas de los enfoques de Campbell y de Scriven son la inferencia a partir de la validez interna de las evaluaciones y la valoración de los resultados mediante comparaciones sumativas de programas alternativos. Cronbach no da prioridad a ninguno de los dos. En contra de Campbell, defiende la prioridad de la validez *externa*; frente a Scriven, aboga por la *evaluación formativa* —la evaluación *dentro* de los programas, y no *entre* programas.

Cronbach es uno de los teóricos de la evaluación más importantes y menos apreciados. Su obra casi nunca se cita en la literatura inglesa sobre evaluación, excepto en algunos apartados importantes de la evaluación educativa (por ejemplo, Simon, 1987), y en la literatura estadounidense, aparte de la importante reseña de Shadish, Cook y Leviton, pocas veces se recurre a su obra, y quienes la utilizan lo hacen de forma selectiva. Sin embargo, la obra de este «tenaz maestro de las distinciones conceptuales» (Scriven, 1986, pág. 15) es

1986a; Simon, 1987; Stake, 1991), y su obra tiene unas amplias implicaciones para aspectos clave de la evaluación cualitativa, por ejemplo:

- la generalización
- las explicaciones causales
- los estudios de casos en enclaves múltiples
- la relevancia política de la evaluación.

ROBERT STAKE: LA EVALUACIÓN RECEPTIVA

Stake fue quizás el primer teórico de la evaluación que situó la indagación cualitativa y los métodos participativos en el primer plano de la tarea evaluadora. Su posición es «compleja, de desarrollo continuado, no extremista» (Shadish y otros, 1990, pág. 468). Su modelo de evaluación receptiva tiene varios elementos:

- Se orienta más directamente a las actividades del programa, más que a los propósitos del mismo.
- Atiende la solicitud de información del interesado local.
- En el informe sobre el éxito o el fracaso del programa se hace referencia a los diferentes valores de las personas.
- Es participativo, con lo que se incrementa el control local.

Stake hace del cambio gradual e incremental una virtud, y tiende a ser políticamente conservador:

Stake no insiste en ayudar a los grupos de interés necesitados, ni en educarles, ni en las causas de su difícil situación. Su planteamiento es el del *laissez-faire*, característico de una perspectiva «derechista» o libertaria (Shadish y otros, 1990, pág. 296).

Stake presta más atención que muchos teóricos a las cuestiones prácticas de la metodología de la evaluación, y quizá sea el principal exponente de los métodos cualitativos del estudio de casos (Stake, 1994, 1995). En la segunda parte de este libro nos extendemos más sobre la evaluación con el estudio de casos. El autor vincula su metodología a un compromiso con un modelo de servicio de la evaluación. La prueba del conocimiento que proporciona la evaluación es su utilidad, con el evaluador que actúa de capacitador y facilitador,

más que como suministrador de ideas. Stake busca la popularización y la democratización del conocimiento. Al hacerlo, se niega a rebasar los enfoques descriptivos de los valores. Los estudios de casos, por lo tanto, se deben «orientar a comprender el caso más o menos como un todo, no para recomponer sus piezas ni fijar sus problemas» (Stake, 1991, pág. 77).

Donde mejor se aprecia su modelo de servicio de la evaluación es en su interpretación de cómo se generaliza lo aprendido a partir de los programas. Al defender unas «generalizaciones naturalistas», Stake y Trumbull dicen:

Creemos que los estudios de evaluación de programas se deben planificar y realizar de forma que ofrezcan toda la experiencia directa posible a los lectores, que así podrán combinarla intuitivamente con su experiencia previa (1982, pág. 2).

Rechazan la «firme expectativa de que el *conocimiento* puede sanar a una educación enferma», como cuando se da por supuesto que «la investigación conduce al conocimiento, que a su vez conduce a una mejor práctica» (1982, pág. 3). Como corolario, estas suposiciones rutinarias ponen el énfasis en la producción, la utilización y la divulgación de los conocimientos. Para Stake esto es un sueño que distrae de la práctica. Él y Trumbull sostienen la idea alternativa de que «la investigación puede provocar una experiencia personal que conduzca a una mejor práctica» (1982, pág. 5).

Una característica importante de la obra de Stake y de otros modelos de evaluación similares ha sido la fertilización mutua de las ideas sobre evaluación que ha tenido lugar entre Estados Unidos y Gran Bretaña. La reforma del currículo en las escuelas inglesas llevó a una evaluación democrática asociada al Centre for Applied Research in Education (CARE) de la Universidad de East Anglia (véase Simon, 1987). Entre los nombres relacionados con CARE están los de McDonald, Parlett, Stenhouse, Elliott, Simon, Hamilton, Walker, Smith, Adelman y Kushner.

La propia postura de Stake sigue avanzando, y el autor ha reflexionado en público sobre tal desarrollo en varias ocasiones (Stake, 1991, 1997). Su extensa revisión de sus propias ideas detalla los avances alcanzados en las posiciones respecto a la metodología, la epistemología y los valores. Su reciente aceptación a regañadientes de las funciones de la evaluación de propugnación o defensa parece marcar un cambio hacia una postura más reformista (Stake, 1997), tal vez como reconoci-

miento tácito de la complacencia política de su obra anterior. En efecto, hace poco se ha dicho de él que piensa que la aparición de las ideas de gestión ha supuesto el fin de la evaluación receptiva (Kushner, 1997).

Los etnógrafos de tendencia general critican la versión que Stake hace de la investigación con estudio de casos y de los modelos de investigación del profesional en la educación (por ejemplo, Atkinson y Delamont, 1993; Hammersley, 1993c), por su ingenuidad metodológica y teórica. No obstante, la obra de Stake es una aportación importante a la evaluación cualitativa en los siguientes sentidos:

- Stake es el primero de los teóricos que hemos examinado que desarrolla un enfoque de la evaluación en el que los profesionales locales son fundamentales.
- Sus iniciativas innovadoras sobre la evaluación participativa han tenido diversos sucesores.
- Su metodología del estudio de casos y su defensa de las generalizaciones naturalistas sientan las bases de posteriores avances.

CAROL WEISS: EVALUACIÓN DE POLÍTICAS

Los estudios de Carol Weiss sobre el contexto político de la investigación y la evaluación de políticas cambiaron, seguramente para siempre, nuestra forma de entender los usos de la investigación evaluativa. La investigación evaluativa de resultados de los años sesenta y principios de los setenta tenía, en el mejor de los casos, un éxito marginal en programas y prácticas innovadoras. Hemos visto que la solución que Campbell dio a esta situación fue desarrollar unos diseños de investigación cada vez más diferenciados para contrarrestar las amenazas a la validez interna. La respuesta neoconservadora cuestionaba la inversión de dinero en los problemas sociales. Más tarde Weiss decía al respecto: «Nos habíamos enrolado como evaluadores con la intención de contribuir a la mejora de los programas sociales, pero parecía que íbamos a terminar por ayudar y tranquilizar a los bárbaros» (citado en Shadish y otros, 1990, pág. 182).

La respuesta de Weiss a las primeras investigaciones de resultados se distinguía tanto de los metodologistas como de los conservadores. Empezó por reconsiderar el lugar de la investigación de evaluación en el diseño de políticas, primero explicando su contexto político, y segundo mediante un trabajo empírico sobre cómo se usa realmente la investigación evaluativa.

En un artículo escrito por primera vez en 1973, la autora sostenía que las consideraciones políticas inciden de tres formas en la evaluación. En primer lugar, los programas y las políticas que se evalúan son ellos mismos criaturas de decisiones políticas:

Surgieron de las batallas violentas y desordenadas del apoyo, la oposición y la negociación políticas. Llevan adjuntas la reputación de los patrocinadores legislativos, las carreras de los administradores, el empleo del personal de los programas y las expectativas de los clientes (Weiss, 1987, pág. 49).

La creación política de programas les priva de claridad, una «consiguiente grandiosidad y confusión» (1987, pág. 51), por lo que suele haber poco acuerdo sobre qué objetivos son reales (en el sentido de que realmente se invierta un esfuerzo para alcanzarlos).

En segundo lugar, la evaluación conlleva unas consideraciones políticas porque sus informes aparecen en el foro político. El trabajo empírico posterior de Weiss se centró en este tema. En tercer lugar, en la investigación de evaluación hay implícita una política. La investigación tiende a aceptar las amplias premisas del programa, y la investigación sería «otorga un aura de legitimidad a la empresa» (1987, pág. 57). La conclusión general de la autora fue defender una investigación más estratégica y lamentar que «la mayoría de estas implicaciones políticas de la investigación de evaluación tienen una orientación del *“establishment”*. Aceptan el mundo tal cual es; tal como se define en la estructura de la organización, en los diagnósticos oficiales de los problemas sociales» (1987, págs. 59-60).

El supuesto racional convencional sobre la utilización de la investigación era que ésta llevaba al conocimiento, el cual, a su vez, proporcionaba una base de carácter instrumental y de ingeniería social para la acción. Las raíces históricas de esta idea son profundas (Finch, 1986). Weiss no fue la primera en cuestionar la legitimidad de esta idea, pero su importancia radica en las razones empíricas, la contundencia de la explicación y la verosimilitud de su exposición. Con sus colegas, Weiss entrevistó a 155 responsables de alto rango de organismos del campo de la salud mental de ámbito federal, estatal y local. Tanto los dirigentes como el personal empleaban la investigación para facilitar información sobre las necesidades del servicio, pruebas sobre lo que funcionaba, y para mantenerse al día en su campo. Pero también se utilizaba como una capa ritual para legitimar posturas y proporcionar una seguridad personal de que la postura asumida era

la correcta. En un nivel conceptual más amplio, ayudaba a los directivos a comprender el mundo. Por todos estos objetivos, «era una fuente entre muchas, y normalmente no lo bastante sólida para dirigir el proceso de toma de decisiones» (Weiss, 1980, pág. 390). En cuanto a la utilización directa, «de hecho parece que el uso instrumental es escaso, sobre todo cuando los temas son complejos, las consecuencias son inciertas, y en el proceso de toma de decisiones, es decir, en el diseño de la política, interviene una multitud de actores» (1980, pág. 397).

El uso de la investigación se reflejaba también en las ideas de los directivos sobre el proceso de toma de decisiones. Se consideraba que éstas estaban fragmentadas en sentido vertical y horizontal dentro de las organizaciones y que eran el resultado de una serie de pasos graduales y amorfos. Por consiguiente, «una destacada razón de que no hablen de un uso de la investigación para tomar decisiones específicas es que muchos de ellos *no creen que tomen decisiones*» (1980, pág. 398). De ahí el título del artículo de Weiss: «La reducción del conocimiento y el acrecentamiento de la decisión». En él se asentaba la tesis de la autora de que la explicación, más que la acción instrumental, representa la ruta característica para el uso de la investigación.

Las ideas de Weiss han sido cuestionadas sobre todo por Patton. La tesis de este autor es doble (Patton, 1988a, 1988b). En primer lugar, argumenta que Weiss se equivoca cuando, a partir de la investigación de políticas, generaliza a la evaluación de programas. «Tiene sentido que la investigación de políticas se emplee de forma más difusa y menos directa que la evaluación de programas», porque son «diferentes tipos y niveles de práctica» (Patton, 1988a, pág. 12). En segundo lugar, la visión de la autora es «bastante sombría». «A mi juicio, la visión de Weiss no es comercializable» (1988a, pág. 11). La respuesta de Weiss es lamentar que en el mundo de Patton «todos se comportan racionalmente» (Weiss, 1988, pág. 18). Además, incide en su argumentación anterior sobre la política implícita en la investigación de evaluación: «La tarea de quien aplique un programa es creer; la del evaluador, dudar» (1988a, pág. 22).

Las teorías de la ilustración o explicación son difíciles de falsar o de verificar, pero parecen verosímiles. Weiss ya no es tan optimista en lo que respecta a los posibles beneficios de la investigación básica: «La ciencia social no es una ruta fácil ni eficiente hacia la reforma social. Produce efectos, pero es probable que sean unos efectos lentos o irregulares» (Weiss, 1987, pág. 67). Además, se siente más optimista sobre las aplicaciones específicas de la investigación: «Se han pro-

ducido mejoras incrementales y, después de todo, no hay que despreciar los incrementos acumulativos» (Weiss, 1987, pág. 67). Hablaremos más de la obra de Weiss en el capítulo 5. Su legado para la evaluación cualitativa es importante y de amplia base. Su obra no está regida por la metodología, aunque casi al final de su carrera publicara un texto de metodología sobre evaluación, del que recientemente ha publicado una nueva edición revisada (Weiss, 1998). Su idea general es que el evaluador que sólo cuenta con la metodología no está bien equipado. Su preferencia personal es el pragmatismo metodológico.

Su enfoque general es directamente coherente con una tendencia importante de la investigación de políticas inglesa. Los estudios y la obra de Abrams, Bulmer y Finch son el exponente de este trabajo, si bien es cierto que los tres insisten más en la necesidad de los métodos de indagación cualitativos y de los usos reformistas de la investigación. Las líneas entrelazadas de sus estudios se unen para abrir un rico filón para la evaluación cualitativa:

Más que por cualquier otra cosa, [Weiss] ha luchado por una teoría realista del uso. Estos cambios iniciaron en la evaluación un debate que sigue vivo sobre el papel del idealismo y el pragmatismo en la evaluación (Shadish y otros, 1990, págs. 207-208).

Los cinco teóricos de la evaluación de que nos ocupamos en la primera mitad de este capítulo muestran diferencias importantes, pero comparten una misma base en lo que se refiere a los términos del debate:

- Todos iniciaron su obra a partir de una fe comprometida en el potencial de la investigación evaluativa para desvelar el mundo real. Campbell y Stake han desarrollado sus ideas de forma que cada vez se acentúa más la incertidumbre epistemológica.
- La preeminencia que los objetivos de los interesados tienen en la obra de Cronbach, Stake y Weiss se ha enmarcado en una política de consenso ampliamente pluralista. Los cinco autores evitan las funciones de defensa.
- Campbell, Cronbach, Weiss y, en menor grado, Scriven centran su obra en los amplios temas de la evaluación de programas y de políticas. La evaluación del profesional surgió únicamente del enfoque de Stake y en el seno del movimiento del profesor como investigador de Inglaterra.

- Los cinco teóricos han adoptado unas versiones más o menos sólidas del pragmatismo metodológico. A excepción de Scriven, todos ellos otorgan un lugar importante a la metodología de investigación tanto cualitativa como cuantitativa. Stake hace de los métodos cualitativos su metodología habitual.

Pero a partir de 1980 la evaluación se ha expandido para reivindicar más un lugar en el mundo, al mismo tiempo que muestra sus dudas respecto a la certeza de los conocimientos que produce. La política de consenso pluralista se ha descompuesto, lo que ha producido un contexto diversificado para la evaluación. House ha desarrollado las tesis de la justicia social, que apoyan una postura reformista. Los teóricos críticos y los grupos feministas han desarrollado unas posturas abiertamente ideológicas, aunque las expresiones de evaluación crítica o feminista en contadas ocasiones afloran en la literatura.

La evaluación del profesional sigue careciendo de un cuerpo teórico extensivo. Sin embargo, los últimos trabajos de Reid en el ámbito de la asistencia social y los de Eisner sobre el especialista y la crítica son ejemplos de avances en este campo. El carácter interpretativo de Stake, Campbell y Eisner se basa en una ontología realista. Se han desarrollado posturas explícitamente relativistas, en especial por parte de Guba y Lincoln, que han cuestionado también el pragmatismo metodológico dominante. En las páginas siguientes, esbozaré los temas fundamentales de cada una de estas convicciones.

ERNEST HOUSE: EVALUAR PARA LA JUSTICIA SOCIAL

La mayoría de los planteamientos de la evaluación anteriores no satisfacen a House por tres razones. Primera, sus ideas de valor son subjetivistas. Tanto si se dirigen a gestores, a élites profesionales o a consumidores, desvinculan la evaluación de las recomendaciones prescriptivas. Segunda, su epistemología es objetivista o intuicionista. Tercera, los supuestos políticos son uniformemente pluralistas. Aunque no les acusa de complacencia política, lamenta que todos se basen en una filosofía esencialmente ilustrada: «Todos presumen que un mayor conocimiento hará que de algún modo las personas sean más felices o se sientan más satisfechas» (House, 1980, pág. 64).

La respuesta mejor desarrollada de House se recoge en su obra principal, *Evaluating With Validity* (1980), e incluye tres aspectos. En primer lugar, sostiene el autor que la lógica de la evaluación no está

tanto en unas pruebas racionales como en la persuasión y la argumentación. Las evaluaciones nunca producen unos conocimientos ciertos: «Sometidas a un serio escrutinio, las evaluaciones siempre parecen equívocas». El hecho de aspirar a un conocimiento cierto «resulta de confundir la racionalidad con la lógica. No son lo mismo [...] Las evaluaciones no pueden ser más que actos de persuasión» (1980, pág. 72). Son actos de argumentación, no de demostración. Y esto es así tanto en la metodología cuantitativa como en la cualitativa, pese a que «las metáforas estadísticas [...] hacen que las pruebas parezcan ciertas e inequívocas» (1980, pág. 74). En resumen: «La evaluación, más que convencer, persuade; más que demostrar, argumenta; más que cierta, es creíble; más que convincente, se acepta de forma variable» (1980, pág. 73).

House habitualmente intercala consideraciones sobre posturas lógicas y morales. Sostiene de forma repetida que se han olvidado las consideraciones normativas y que «si esto es un defecto en la conducta de la ciencia, en la evaluación es un error fatal» (1980, pág. 251).

En segundo lugar, House ha desarrollado una amplia argumentación en favor de un propósito reformista y de justicia social de la evaluación. No es indiferente ante la estética de la evaluación. Sin embargo, para él, «la verdad es más importante que la belleza. Y la justicia, más importante que una y otra» (1980, pág. 117). Las teorías utilitaristas de la justicia, basadas en la mayor satisfacción global, ofrecen unos criterios claros para la evaluación, pero tienden a favorecer a las clases altas, llevan a unos juicios que «no se concilian con la sensibilidad moral de la persona» (1980, pág. 134) y a veces conducen a una simplificación exagerada. La teoría pluralista e intuicionista se acerca más al sentido común, a los juicios cotidianos de justicia, y fomenta que se insista en la representación de las opiniones de los interesados. Pero «la amenaza del relativismo» (1980, pág. 134) pone en peligro la coherencia de la aplicación y tiende a otorgar un valor excesivo al juicio profesional.

En esencia, el modelo pluralista confunde las cuestiones de intereses con los conflictos de poder. Sólo puede equilibrar aquellos intereses que se representan, normalmente los de los poderosos (House, 1991a, pág. 240).

El concepto de justicia de House encuentra su aplicación más práctica en sus detallados argumentos en favor de un acuerdo de evaluación justo (1980, cap. 8) y en sus tesis más amplias sobre la ética

de la evaluación (1993). La sustitución que hace de las ideas tradicionales de objetividad por la de imparcialidad delimita su postura reformista explícita. En respuesta a la críticas de las corrientes más generales, que dicen que su acuerdo de evaluación justo es tendencioso a favor de los desfavorecidos, replica:

Me parece que asegurarse de que se representan y se consideran seriamente los intereses de los desfavorecidos no significa ser tendencioso, aunque sin duda es una actitud más igualitaria que la de muchas de las prácticas actuales (1991a, págs. 241-242).

En tercer lugar, el autor pretende resolver los problemas epistemológicos de los anteriores enfoques de la evaluación a través de un complicado realismo científico, del que espera que pueda servir de base para una nueva síntesis que resuelva la «guerra de paradigmas» que se libra dentro de la evaluación (House, 1991b). Volveremos a este tema en el capítulo 3, y aquí sólo señalamos que el realismo científico implica una forma de entender la «causa» y la «objetividad» muy distinta de lo que son las ideas convencionales. El autor pretende también integrar la lógica evaluativa y la justicia, con el fin de desarrollar una base para sintetizar los resultados de la evaluación (House, 1995). En el capítulo 9 nos ocupamos de estas tesis.

Al igual que Stake, House expone sus primeros trabajos a la influencia fertilizante de los desarrollos producidos en Gran Bretaña (Simon, 1987), aunque su obra posterior se ha estudiado menos fuera de Estados Unidos. Su planteamiento de los temas relacionados con la justicia tiene unas lagunas que él mismo reconoce. Su obra principal se concluyó cuando aún estaban apareciendo las aplicaciones neomarxistas de la justicia y antes de que se emprendieran de forma constante los estudios feministas. Es muy consciente de que los supuestos subyacentes de estas posturas son distintos tanto de los suyos como de los de sus predecesores teóricos. Tiende a presumir claramente una idea abiertamente consensual de la sociedad, que limita su postura reformista. Piensa que, aunque los pensadores, los legisladores y el público disientan, lo hacen «en el marco de un acuerdo general sobre los valores democráticos fundamentales» (House, 1995, pág. 44). Su entrega a los modelos de evaluación participativos está limitada por el mantenimiento que hace de la función del evaluador experto.

Los planes que esboza para la evaluación cualitativa son extensivos. Su idea de justicia es sutil, tanto cuando advierte de que la práctica de la evaluación no está *determinada* por nuestra idea de justicia,

como cuando distingue la relatividad cultural y moral al aceptar la adecuada diversidad de los conceptos de justicia. En su conjunto, las obras de House, Cronbach y Campbell constituyen una sólida base para las inferencias causales mediante la evaluación cualitativa. Los programas no son entidades fijas que actúen de «estatua» (en expresión de Cronbach), sino que varían de un sitio a otro. Las causas se descubren mediante el conocimiento de los detalles —«aquellas evaluaciones que prevén y siguen la variabilidad y la irregularidad de los acontecimientos» (House, 1991b, pág. 8)—. En estas circunstancias, opta por la metodología cualitativa. El trabajo de House ha sido coherente con sus propias expectativas de «un orden conceptual más completo y más transparente [...] un sentido más fuerte de responsabilidad moral» (1980, pág. 11).

LA EVALUACIÓN CRÍTICA

Según la idea reformista de House, los desacuerdos entre los grupos de interés de la evaluación se producen dentro de un consenso general en lo que se refiere a los valores democráticos fundamentales. House observó el surgimiento de los modelos críticos de evaluación, y reconoció que cualquier persuasión de este tipo referente a la evaluación quedaría excluida de su esquema para la comprensión de la ubicación política de la evaluación.

Veinte años después, todavía no existe ningún teórico destacado de la «evaluación crítica» al que se pueda considerar el «representante» de la aplicación de la teoría crítica a la evaluación. No es que no exista un arduo debate sobre la investigación crítica (Carr y Kemmis, 1986; Harvey, 1990) y versiones neomarxistas (Willis, 1977), feministas (Acker y otros, 1983; Humphries, 1999), antirracistas (Whitmore, 1994), freirianas, participativas (Edwards, 1989; Martin, 1996), anti-discapacitantes (Oliver, 1992), orientadas a la praxis (Lather, 1986b), deconstructivistas (Lather, 1991) y de potenciación (Dullea y Mullender, 1999; Everitt y Hardiker, 1996; Miller, 1990) de una visión liberadora de la evaluación. Todas ellas tienen unas consecuencias implícitas para la evaluación cualitativa, aunque la mayoría de estos profesionales y autores hablarían de «investigación», más que de «evaluación», para referirse a sus teorías.

La evaluación crítica es «crítica» en el sentido de que los problemas de la evaluación se conceptualizan dentro de su contexto social, político y cultural. Es una «evaluación» crítica, para emplear las pa-

labras de un excelente trabajo de Popkewitz sobre el paradigma de la teoría crítica, porque «se refiere a una indagación sistemática que se centra en las contradicciones de [...] la práctica» (1990, pág. 46). El teórico crítico reconstruye las reglas de la evaluación para transmitir la idea de que las categorías lingüísticas de la investigación y de la evaluación están relacionadas históricamente con las cuestiones sociales y morales más amplias de la producción y la reproducción: «La metodología se ocupa del orden moral [...] presupuesto en la práctica de la ciencia. Es el estudio de lo que se define como conocimiento legítimo» (1990, págs. 51-52). Implica un análisis de base teórica que se distingue por un escepticismo ante las fuerzas de la regulación y la distribución sociales y, a veces, también ante las pruebas empíricas: «Se trate de lo que se trate, no hay que aceptarlo sin más» (Harvey, 1990, pág. 8). Los evaluadores críticos rechazan la idea de la ciencia como un desarrollo acumulativo y progresivo. Generalmente son realistas falibles en su concepción filosófica y pueden resultar menos inflexibles que autores constructivistas como Guba y Lincoln, en la medida en que puede existir un acuerdo metodológico entre quienes defienden diferentes paradigmas. Por ejemplo, Lather se niega a demonizar otras convicciones e insiste en que «no existen posturas inocentes» (1991, pág. 85).

Existe un debate entre los científicos críticos sobre hasta qué punto participan en un proyecto de carácter emancipador y liberador. Algunos, por ejemplo quienes defienden una postura feminista, piensan que el papel partidista de la ciencia les obliga a buscar compromisos políticos a través de «una implicación directa y explícita en los esfuerzos por transformar las actuales relaciones sociales» (Popkewitz, 1990, pág. 49). Se entiende que la investigación está orientada a la praxis (Lather, 1986b), donde «un proceso recíprocamente educativo es más importante que el producto» (Lather, 1991, pág. 72). Otros, incluido Popkewitz, no están de acuerdo, y piensan que «los científicos sociales son partidistas cuando elaboran unas agendas sociales mediante las prácticas de la ciencia» (1990, pág. 50).

Más que construir una agenda para la evaluación cualitativa a partir de una evaluación crítica latente, prefiero indicar en qué términos se puede desarrollar una crítica de otras formas de evaluación. En mi opinión, la mayoría de los evaluadores críticos se *opondrían* a la mayor parte de lo siguiente:

- Una distinción entre evaluación e investigación basada en que la primera no implica *teorizar*.

- Considerar que las posturas paradigmáticas sólo representan un conflicto de *ideas*.
- Una evaluación basada en la presunción de un *consenso* democrático.
- Pensar que la evaluación sólo tiene sentido dentro del *contexto local* de la indagación.
- Entender la evaluación como esencialmente asociada con una *toma de decisiones planificada*. Se podría recordar la paradoja de Marx de que cualquiera que haga planes para después de la revolución es un contrarrevolucionario.
- Un enfoque de *valoración descriptiva* de las conclusiones evaluativas.
- Una presunción *gradualista* e incremental sobre los cambios deseables que deriven de la evaluación.
- Un contrato de evaluación *regido por la dirección*.
- Una evaluación *participativa ingenua* que no reconoce que «la intención emancipadora no garantiza un resultado emancipador» (Acker y otros, 1983, pág. 431).
- Una *ortodoxia metodológica*, sea de carácter cuantitativo o evaluativo. Lather, por ejemplo, critica la investigación que no consigue «ir más allá del simple uso de los métodos cualitativos» (Lather, 1992, pág. 91).

WILLIAM REID: EVALUACIÓN PARA DESARROLLAR LA PRÁCTICA

Reid ha desarrollado su carrera profesional en la enseñanza de la asistencia social. Sus primeros trabajos surgieron de lo que era la corriente general de la evaluación experimental de resultados. Su obra más reciente se distingue de la de los teóricos que hemos considerado en que este autor basa el desarrollo y la evaluación continuados de los modelos orientados a los objetivos en una dedicación creciente a la investigación cualitativa y orientada a los procesos. Sus ideas agudizan los temas de la evaluación cualitativa; los extienden al tiempo que los cuestionan.

En su obra *Brief and Extended Casework* (Reid y Shyne, 1969), se expone un estudio experimental en el que se compara la intervención de asistencia social a corto y largo plazo. La provisión planificada de servicios a corto plazo produjo unos resultados al menos tan buenos como la intervención sin plazos definidos, y posiblemente mejores. En el contexto del tiempo, cuando se consideraba que el trabajo de

asistencia social individual era el ideal, «en esta carrera concreta (la Intervención Planificada a Corto Plazo) sólo necesitaba un empate para ganar» (1969, pág. 175).

Reid fue el primero en desarrollar y comprobar la Práctica Centrada en Tareas, un planteamiento del trabajo social basado en la solución de problemas a corto plazo, cuyo objetivo está en los problemas que los clientes consideran que son objeto de preocupación. El modelo se ha documentado exhaustivamente con investigaciones y tiene una fuerte orientación empírica (Reid, 1997). Este hilo conductor que recorre la carrera de Reid ofrece el contexto de toda una diversidad de preocupaciones relacionadas con la evaluación.

En primer lugar, sus propios estudios de resultados y su obra de reseña de otros estudios han sido un puntal en la gradual renovación de la confianza en este tipo de investigación evaluativa (Reid, 1988, 1997; Reid y Hanrahan, 1982). En sus trabajos se señala «el dominio de unas formas de práctica estructuradas», en particular las que están influidas por «la conducta del movimiento de modificación» (Reid y Hanrahan, 1982, pág. 329). Está convencido de que las profesiones asistenciales pueden demostrar que son eficaces, pero advierte de que la relevancia práctica de los efectos identificados muchas veces es escasa, y la duración de éstos, ambigua. A veces parece que duda de los criterios metodológicos de este tipo de estudios. En su aleccionadora y memorable imagen, «es como intentar decidir qué caballo ha sido el ganador, cuando se ha visto la carrera desde un mal ángulo de tribuna y en medio de un chaparrón» (Reid, 1988, pág. 48).

En sus observaciones sobre la dirección que sigue la investigación de resultados, dice en sentido favorable: «El investigador como evaluador está dando paso al investigador como persona que desarrolla modelos» (Reid y Hanrahan, 1982, pág. 338). Es un reflejo del que es otro de sus intereses referentes a la evaluación. Entiende su propuesta de una investigación de la práctica centrada en las tareas como un paulatino proceso de desarrollo, comprobación y posterior desarrollo. Ha intentado desarrollar para el campo unos métodos discretos y agradables para el profesional, que comprueban las primeras versiones experimentales de las intervenciones específicas en la práctica (Reid, 1994b).

La modesta confianza de Reid en los beneficios de la investigación evolutiva y de resultados contrasta con otros pronunciamientos más optimistas referentes a la práctica basada en la experiencia. La respuesta de Reid a estos planteamientos ha sido una revisión modera-

da del movimiento de la práctica empírica (Reid, 1994c; Reid y Zettergren, 1999):

El movimiento de la práctica empírica ha fomentado la antigua y laboriosa pretensión de crear una profesión de base científica. Sin embargo, las ganancias obtenidas han sido parciales, desiguales y no siempre claras (Reid, 1994c, pág. 180).

Estas palabras reflejan su tercera aportación a la evaluación, cuando sugiere de qué forma se deben aplicar a la asistencia social las consideraciones sobre epistemología y metodología. Por ejemplo, ha explorado el potencial de la «investigación del proceso de cambio» (Reid, 1990). En su opinión, las limitaciones prácticas y éticas de los experimentos hacen necesario un estudio naturalista de las relaciones entre proceso y resultados. Una idea clave sobre el proceso de cambio es la de los «microrresultados» intermedios que se muestran mediante el estudio de elementos más reducidos de la práctica. De este modo, se da una importancia capital al contexto inmediato de la práctica. Este sistema de investigación tiene el potencial de que se puede orientar hacia una metodología naturalista e interpretativa más sólida, o hacia otras ideas más tradicionales del rigor. Reid se opone a esta última, y probablemente a la primera, y prefiere pensar que el conocimiento se alcanza como «una especie de conocimiento provisionalmente válido» (1990, pág. 144). En el fondo de esta cuidada orientación está la preocupación de Reid por reenmarcar el debate prescindiendo «de las discusiones sobre visiones epistemológicas del mundo y, en su lugar, emplear nuestras energías en esclarecer y resolver las cuestiones concretas presentes en estas discusiones» (1994a, pág. 465).

Su interés está centrado en la buena práctica, el trabajo empírico, un vocabulario y una racionalidad compartidos, y una ausencia de entusiasmo por la autoridad filosófica. Pertenecer «a la tradición del pragmatismo norteamericano, especialmente tal como se ha expresado en las obras de Peirce, James, Dewey y Rorty» (Reid, 1994a, pág. 472). También desea «redefinir la naturaleza de las tendencias generales, para que la metodología sea parte, y no se aparte, de ella» (1994a, pág. 477).

Su obra plantea unas cuestiones inevitables que no se han resuelto aún. En su aceptación de la experimentación incremental y en sus comentarios existe el supuesto gradualista de que «estos pequeños éxitos de hoy son mejores que los grandes fracasos del pasado» (Reid y Hanrahan, 1982, pág. 338). Reid tiende a ignorar otras cuestiones

más amplias, como la de la política implícita en la investigación de evaluación. Es consciente de que si los trabajadores sociales centran su energía en la consecución de unos resultados modestos pero eficaces, es posible que «se frenen otras innovaciones más radicales, aunque posiblemente menos comprobables» (Reid, 1988, pág. 45).

Reid conoce mejor que la mayoría de los teóricos de la asistencia social la literatura más general sobre evaluación, aunque el proceso de fertilización entre las distintas obras en su caso sigue siendo limitado. Es una lástima. El compromiso de Reid con la investigación evolutiva significa una ampliación imaginativa del alcance de la evaluación. Sus trabajos sobre la investigación del proceso de cambio ilustran cómo ha sabido «adaptar, de forma pragmática y efectiva, una serie de ideas y de procedimientos de su forma de investigación orientada al proceso a su modelo de práctica orientada al objetivo» (Sherman, 1990, pág. 152). Además, ha insinuado de qué forma se puede combinar la investigación del proceso de cambio con estrategias de investigación evolutiva. Su obra reciente sobre la práctica empírica como perspectiva supone un reto interesante a las posturas paradigmáticas más arraigadas (Reid y Zettergren, 1999). La investigación evolutiva y la adaptación que Reid hace de los análisis de resultados basados en el proceso de cambio mejorarán «la precisión de nuestro propio desconcierto» (Geertz, 1973, pág. 29).

ELLIOT EISNER: LA EVALUACIÓN REFLEXIVA

Al igual que Reid, Eisner centra la indagación evaluativa en la tarea profesional inmediata. Pero aquí terminan más o menos sus coincidencias. A Eisner le interesa la enseñanza, en especial la pedagogía. Su objetivo es que los profesores lleguen a ser «a la vez quienes perciban y quienes comuniquen lo percibido» (Eisner, 1991a, pág. 3), la experiencia y su representación. La más completa exposición de sus ideas se encuentra en *The Enlightened Eye** (1991a). La experiencia conlleva algo más que la simple observación: «La experiencia es una forma de consecución humana [...] *aprendemos* a ver, a oír y a sentir» (1991a, pág. 21). La comunicación de esta experiencia a otras personas es fundamental en sus planteamientos: «El escritor empieza por la experiencia y termina en las palabras. El lector empieza por las palabras y termina en la experiencia» (1991a, pág. 22).

* *El ojo ilustrado*, Barcelona, Paidós, 1998.

En el centro del modelo de Eisner hay dos ideas: la especialización educativa (la del entendido en educación, para tomar una expresión propia de la enología o del arte) y la crítica educativa. La especialización es el arte de la apreciación, y conlleva una «inteligencia cualitativa» (1991a, pág. 64) que al profesional le permite superar las «percepciones de reserva» y el problema de que «una forma de ver es también una forma de no ver» (1991a, pág. 67). Habla de «observación epistémica» para referirse a una apreciación que se puede basar en los cinco sentidos. Esencial en el entendido parece ser la reflexividad, aunque Eisner hable de ella sólo de forma ocasional:

Cuando actuamos como entendidos, es importante que centremos la atención en dos objetivos: uno son los propios acontecimientos; el otro es lo que esos acontecimientos provocan en nuestra experiencia (Eisner, 1991a, pág. 189).

Su segunda idea, la crítica educativa, es el arte de la revelación: «Los entendidos no necesitan más que apreciar aquello con que se encuentran. Los críticos, sin embargo, deben dar vida a estas cualidades mediante un uso ingenioso de la revelación crítica» (Eisner, 1988, pág. 143). La especialización es un acto privado: «La crítica da a la especialización un rostro público» (Eisner, 1991a, pág. 85). Aunque es posible ser entendido sin ser crítico, «no se puede ser crítico de ningún tipo sin cierto grado de especialización» (1991a, pág. 86). En general, presume que una misma persona ejerce la especialización y la crítica.

Sus ideas sobre cómo tiene lugar la generalización del aprendizaje tienen su valor, y además ha desarrollado también unos test de criterios para la práctica ética. En un sentido general recomienda los valores, y al crítico le exige que juzgue el valor educativo de lo que ocurre. Según Eisner, esto es lo que distingue a la crítica educativa de la etnografía. Pero en casos concretos es un valorador descriptivo: «Creo que la principal aportación de la evaluación es una mayor conciencia de las cualidades de esa vida, con el fin de que profesores y alumnos puedan llegar a ser más inteligentes en ella» (Eisner, 1988, pág. 142).

Se ha dicho acertadamente que las ideas de Eisner son «audaces e imaginativas» (House, 1980, pág. 235), pero siguen siendo incompletas y sólo están desarrolladas en parte. Su metodología es inequívocamente cualitativa. No excluye en ésta la existencia de *elementos* cuantitativos, e insiste en que «en la indagación cualitativa *caben las*

cifras. Y lo repito: caben las cifras» (1991a, pág. 186). La diferencia entre cuantitativo y cualitativo es en cierto modo similar a la que expresa uno de los aforismos favoritos de Dewey: «La ciencia establece hechos, el arte los expresa». Sin embargo, en lo que se refiere a los métodos concretos se distingue por su vaguedad: «No conozco ningún “método” para la indagación cualitativa en general, ni para la crítica educativa en particular». Y: «Todo lo que sirva para ver con más exactitud y comprender con mayor profundidad es juego limpio» (1991a, págs. 169, 82). No habla de análisis, y se ha señalado que «los peligros de pasar directamente de los datos al relato nos parecen algo real y grave para la crítica educativa» (Pitman y Maxwell, 1992, pág. 750).

Pese a sus exposiciones recurrentes de los conceptos definitorios, la descripción que se hace del proceso sigue siendo muy vaga. ¿El entendido es el profesor, u otro investigador? Habla de las dos posibilidades, pero las implicaciones metodológicas son distintas en cada caso. ¿Corre el peligro de exagerar la exclusividad de la buena enseñanza y de negar los elementos rutinarios e incluso formulistas de cierta buena práctica profesional? ¿Realmente funciona la metáfora de saber apreciar un buen vino o un buen cuadro? Los vinos y los cuadros no someten a un interrogatorio a quien los percibe en el mismo sentido en que los niños pueden someter a un interrogatorio a los asistentes sociales, los profesores o los médicos (véase Greene, 1992, para comentarios comparables sobre fuertes énfasis estéticos de la evaluación constructivista). En efecto, las metáforas básicas no ayudan en nada al aparente carácter no participativo del tipo de indagación evaluativa de Eisner.

Eisner admite que en cierto sentido las implicaciones de sus ideas están «plagadas de optimismo» (1990, pág. 101). Pero, a pesar de estas críticas, el modelo de Eisner establece una agenda clara e importante para la evaluación cualitativa. Su bien elaborada defensa de la metodología cualitativa y sus distinciones entre *investigación* cualitativa, *indagación* cualitativa y *razonamiento* cualitativo apuntan a un solapamiento de fronteras tanto en la investigación como en las actividades de la vida cotidiana. Su planteamiento tiene puntos de contacto con otros autores que se ocupan de la indagación reflexiva, por ejemplo Schon, Reason, England, Heron, Gould, Fook y Shaw, de algunos de los cuales hablamos en el capítulo 6. Al igual que Stake y Reid, Eisner tiene unos planes prácticos para la mejora de la práctica local, y los detalles de su obra, por ejemplo sobre la ética, son, en consecuencia, más verosímiles.

Como la mayoría de los teóricos de quienes hemos hablado, Eisner empezó sus trabajos en el contexto de la educación. Sin embargo, y a diferencia de otros, no se ha alejado de sus intereses educativos. Tal vez corresponda a otros explorar la posibilidad de trasladar sus ideas generales fundamentales a otros contextos profesionales, como los de la salud, la asistencia social o el derecho.

EGON GUBA E YVONNA LINCOLN: LA EVALUACIÓN CONSTRUCTIVISTA

Stake introdujo una tendencia interpretativa en la teoría de la evolución. En efecto, dice: «He ido sustituyendo la presunción realista por la duda constructivista» (Stake, 1991, pág. 81). Pero si la epistemología de Stake sigue siendo ambigua, el paradigma relativista, constructivista e infundido de valores de Guba y Lincoln ha cambiado nuestra forma de pensar respecto a la evaluación.

La crítica que hacen a la indagación tradicional es bien conocida. Como la resume Lincoln en su excelente exposición personal de su viaje hacia el constructivismo: «Egon y yo rechazamos la indagación convencional por tres razones fundamentales: su postura sobre la realidad; su idea de la relación entre quien conoce y lo conocido; y su posición respecto a la posibilidad de generalizar» (1990, pág. 68).

Respecto a la realidad, abogan por unas realidades múltiples y construidas socialmente que «cuando se conocen con mayor extensión, tienden a producir una investigación divergente» (Lincoln y Guba, 1986a, pág. 75). Las realidades no se pueden estudiar «por partes» (por ejemplo, como variables), sino sólo de forma holística y en contexto. Se rechaza la imagen tradicional de las relaciones entre el conocedor y lo conocido, el investigador y el objeto de la investigación: «El conocedor y lo conocido no sólo no *pueden* permanecer distanciados y separados en el proceso de la evaluación, sino que probablemente *no deben*» (Lincoln, 1990, pág. 68). Por consiguiente: «La relación, cuando se establece adecuadamente, es una relación de negociación respetuosa, de control conjunto y de aprendizaje recíproco» (Lincoln y Guba, 1986a, pág. 75). Por último, como no existen afirmaciones sobre la verdad que sean duraderas y libres de cualquier contexto y dado que la conducta humana está vinculada al tiempo y al contexto, «empezamos a dudar seriamente de la posibilidad de generalizar de una situación a otra» (Lincoln, 1990, pág. 68).

El relativismo es quizá la idea más importante de su modelo de «evaluación de cuarta generación» (Guba y Lincoln, 1989). Guba pro-

voca intencionadamente cuando subraya: «La única alternativa al relativismo es el absolutismo» (Guba, 1990, pág. 18). El relativismo y el constructivismo son coherentes. Para Guba y Lincoln no existe más realidad que la que crean las personas cuando intentan comprender lo que les rodea y la evaluación crea la realidad que expone, no descubre ninguna realidad que exista objetivamente. Un informe es simplemente el «residuo» que deja el proceso que creó los «hechos». La metodología de la evaluación constructivista actúa mediante la hermenéutica (provocar y perfeccionar las construcciones) y la dialéctica (comparar y contrastar, y enfrentarse a las construcciones de otras personas). El objetivo es alcanzar «unas construcciones cada vez más informadas y más complejas» (Guba, 1990, pág. 27).

La obra de Lincoln y Guba es igualmente importante por el análisis que hacen de otros enfoques desde el punto de vista de los paradigmas, o los conjuntos fundamentales de creencias que orientan la acción. Su análisis de los paradigmas parte de su crítica de la naturaleza de la realidad (ontología), las relaciones entre el conocedor y lo conocido (epistemología) y la metodología. Los paradigmas se han presentado con diversos términos, por ejemplo, modernos/posmodernos y cuantitativos/cualitativos. Guba y Lincoln distinguen entre positivismo, pospositivismo, teoría crítica y constructivismo, aunque, como la mayoría de los teóricos críticos, tienden a considerar el pospositivismo como «una versión modificada del positivismo» (1990, pág. 20). La adopción de un paradigma «literalmente impregna todos los actos, incluso los que sólo tienen una relación tangencial con la investigación» (Lincoln, 1990, pág. 81).

Han dedicado gran parte de sus esfuerzos a desarrollar unos criterios de validez que resulten apropiados para la indagación naturalista y constructivista. Esto ha supuesto una serie de criterios de «veracidad» análogos a los criterios de la investigación tradicional. Posteriormente, desarrollaron unos «criterios naturalistas intrínsecos» de la «autenticidad» que principalmente se refieren a cuestiones éticas e ideológicas (Lincoln y Guba, 1986a).

Además de estos trabajos, han considerado también las diferencias de metodología entre la investigación convencional y la naturalista (Guba y Lincoln, 1989); la relación entre evaluación e investigación (Lincoln y Guba, 1986b); y las cuestiones éticas que plantea la investigación naturalista y positivista (Lincoln y Guba, 1989).

El trabajo de estos teóricos es de gran importancia, e influye en la mayor parte de los temas que analizamos en los dos capítulos siguientes. De momento, baste con señalar algunas de las cuestiones

más inmediatas que suscita su obra. Algunos comentaristas, en general favorables, critican a Guba y Lincoln que no sean lo bastante coherentes con sus principios y que conserven las ideas convencionales de unos sujetos autónomos y racionales (Schwandt, 1997). También se ha hablado mucho de los paradigmas y de su relación tanto con la teoría como con la práctica de la evaluación. Pitman y Maxwell concluyen que

la mayor virtud de la obra de Guba y Lincoln se encuentra en la crítica de otros enfoques; su mayor defecto, en el modelo para realizar activamente una evaluación con criterios naturalistas o constructivistas (1992, pág. 742).

La crítica no es del todo justa. Las estrategias de Guba y Lincoln para la validez consensual están orientadas directamente a la práctica. Una crítica más seria es la que se refiere a la presunta incoherencia entre la retórica y la práctica evaluativa de los autores (Heap, 1995). En el siguiente capítulo me ocupo de esta crítica. Guba y Lincoln insisten en que los valores son inherentes al investigador, al paradigma de evaluación, a las decisiones respecto a la teoría y al contexto de la evaluación, y en que esta evaluación potencia a los interesados. Sin embargo, no estoy seguro de si su postura respecto a los valores se extiende más allá de la idea pluralista y potencialmente conservadora de que «todas las ideologías deben tener la misma oportunidad de expresión en el proceso de negociación de las recomendaciones» (Lincoln y Guba, 1986a, pág. 79). Si, como ellos dicen, «lo que hay que cambiar es la mente, no el mundo real» (Guba, 1990, pág. 27), el potencial de cambio reformista o crítico puede ser, en el mejor de los casos, indirecto y contingente. Aunque no hay duda de que el paradigma de Guba y Lincoln ha cambiado nuestras ideas sobre la evaluación, necesitamos unos ejemplos de su aplicación para desarrollar y juzgar su validez.

CONCLUSIÓN: AGENDA PRELIMINAR PARA LA EVALUACIÓN CUALITATIVA

Concluiré este capítulo con un inventario provisional de preguntas que plantea la lectura de los teóricos aquí presentados.

- ¿El razonamiento evaluativo es similar al razonamiento de investigación, o es diferente?

- ¿La diversidad de paradigmas es útil o no en el desarrollo de una evaluación cualitativa?
- ¿Qué implicaciones tienen para la evaluación cualitativa los desarrollos de las versiones modificadas del realismo?
- ¿Qué consecuencias se derivan para la evaluación cualitativa de la comprensión del contexto político en que se encarga la evaluación y en el que se informa?
- ¿La evaluación cualitativa puede abordar problemas de inferencia causal?
- ¿De qué forma, si es que existe alguna, se puede generalizar a partir de las situaciones en que se realiza la evaluación?
- ¿Cuáles son los problemas que plantean la evaluación cualitativa *de programas* y la evaluación cualitativa *del profesional*?
- ¿Los métodos de la *evaluación* cualitativa son diferentes de los métodos de la *investigación* cualitativa?
- ¿Qué potencial tienen los planteamientos de diseño y desarrollo para las aplicaciones cualitativas?
- ¿Cuál es la mejor práctica que podemos encontrar hoy para la evaluación participativa?
- ¿Qué problemas y qué principios éticos se suscitan en las fases de diseño, recogida de datos, análisis e informe de la evaluación cualitativa?
- ¿Los evaluadores cualitativos deben adoptar una postura descriptiva o prescriptiva en lo que se refiere a la valoración?
- ¿En qué circunstancias puede ser necesaria una postura partisana o emancipadora en la evaluación cualitativa?
- ¿De qué forma *se utiliza y se debería utilizar* la evaluación cualitativa?
- ¿Cómo se puede mantener una postura de autocrítica reflexiva en la evaluación cualitativa?

No presumo que sean unas preguntas coherentes. En el capítulo no se ofrece ni una taxonomía ni una historia, y estas preguntas tan sólo reflejan los temas que han preocupado a los teóricos de la evaluación. Y mucho menos pretendo ser exhaustivo. Por ejemplo, en la lista se habla poco de los métodos en general, o de los métodos de evaluación en particular. De hecho, para la mayoría de estos autores, las cuestiones relacionadas con los métodos no han sido fundamentales para la práctica de la indagación. Por último, espero que no haya transmitido en exceso las que son mis predisposiciones. Sin duda, podría desear que hubiera sido menor el pragmatismo metodológico

en la mayor parte de las teorías sobre la evaluación, como también podría desear que hubiera más, o *alguna*, fertilización mutua entre los evaluadores de las diferentes disciplinas y profesiones. El problema lo admite Cronbach, y sólo es abordado, de forma un tanto particular, por Scriven. En el capítulo siguiente examino unas amplias alternativas: paradigmas o pragmatismo, realismo o relativismo, y rigor o relevancia.

CAPÍTULO 3

LAS PRUEBAS OBTENIDAS CON LA EVALUACIÓN CUALITATIVA

Los profesionales actúan o declinan actuar, se sienten seguros o inseguros de contar con una base sólida para sus actos y sus decisiones, establecen unas prescripciones o se declaran agnósticos, y todo ello a la luz (o en la penumbra) de las mejores pruebas que razonablemente puedan reunir. ¿Pero cuál es el estatus de estas pruebas? ¿Pueden los profesionales basar de forma segura sus actos en el supuesto de que éstas ofrecen una aproximación lo suficientemente buena a la realidad, que se podrá trasladar a situaciones nuevas similares? ¿O deben seguir con su profundo escepticismo sobre toda información similar dado que ésta es contingente, un producto relativo de un contexto local exclusivo y que puede representar también los intereses y las construcciones de los poderosos?

Se defiende hoy con energía la necesidad de conseguir que las pruebas «funcionen» para la práctica profesional. La propia expresión de una «práctica basada en pruebas» supone una beligerante seguridad, el tono de «hay que demostrarlo». El lenguaje que habla de metas, objetivos, resultados y eficiencia cuestiona la confianza en los sentimientos, la práctica basada en la opinión, la intuición o el conocimiento no especializado. La expresión, junto con la pregunta que le acompaña: «¿Qué es lo que funciona?», se ha ido introduciendo en las formulaciones oficiales. Se ha utilizado en la literatura académica sobre investigación de evaluación al menos desde 1980 (Weiss, 1980, pág. 387) y era habitual en los informes oficiales en Estados Unidos hacia mediados de los años ochenta (US Department of Education, 1986). Aunque se suele considerar que la medicina es el modelo que los aspirantes deben emular (Sackett y Rosenberg, 1995), los criterios basados en pruebas se emplean para evaluar la enseñanza, la atención a personas mayores, la asistencia social, la política de juventud, la policía, las prisiones, la farmacia, las bibliotecas y las prácticas de empresa (Kirkpatrick y Lucio, 1995).

En el capítulo 1 apuntábamos brevemente las razones de por qué la práctica basada en pruebas no ha de ser el primer recurso que deban emplear los profesionales o los evaluadores que deseen establecer unas bases seguras para la práctica eficaz. Sin embargo, la idea de práctica basada en las pruebas tiene la virtud de cuestionar el concepto de prueba en dos sentidos. En primer lugar, la práctica basada en pruebas ofrece un útil respaldo al importante supuesto de que algunos temas fundamentales referentes a las pruebas no son especiales de ninguna profesión, sino comunes a trabajos superficialmente muy distintos. La práctica basada en pruebas fomenta esta suposición al formular con insistencia las mismas preguntas básicas referidas a todos sin excepción. La investigación de base cualitativa también fomenta el mismo supuesto a través de un planteamiento más inductivo y de procesos inversos, de abajo arriba, para comprender la práctica, despojado en parte de anteriores suposiciones a propósito de que ya se sabe en qué consiste realmente una determinada profesión. En segundo lugar, la práctica basada en pruebas dirige la atención hacia la práctica directa y no hacia estrategias y agendas de evaluación de programas, que han dominado las teorías de la evaluación.

En este capítulo se analizan los temas de fondo que configuran el debate y la práctica referentes a las pruebas. Se plantean, quizá de forma excesivamente simple, como una serie de alternativas:

- ¿La evaluación se debe basar en un *paradigma* concreto de la práctica evaluativa, o en iniciativas *pragmáticas* e incluso oportunistas?
- ¿Se puede descubrir una imagen objetiva y *real* del mundo, o todas las explicaciones evaluativas son *relatividades* contingentes?
- ¿Hay que planificar la evaluación según las exigencias del *rigor* científico, o según la *relevancia* práctica?

En el capítulo 4, nos situamos en otras bases para considerar que las respuestas a estas preguntas han de tener también en cuenta una serie de cuestiones más «orientadas al uso».

- ¿Cómo se sabe si se hace bien la evaluación?
- ¿Los resultados de la investigación evaluativa se emplean mediante la aplicación directa a los problemas, o mediante un proceso de explicación más general?
- ¿Qué principios éticos y qué reglas de decisión deben guiar la evaluación cualitativa?

- ¿Cómo deben configurar las respuestas que se den a estas preguntas la metodología que se utilice?

PARADIGMAS Y PRAGMATISMO

«Paradigma» es una palabra espinosa. En efecto, se ha convertido en una «palabra cubo» (Popper, 1966, 1989) que contiene significados diversos. Si se toma en el sentido general de «una serie básica de creencias que guían la acción» (Guba, 1990, pág. 17), se avanza muy poco. Hasta para el filósofo menos tendencioso supondrían toda una empresa cada una de las cinco palabras clave de esta definición. ¿Cuántas «series básicas de creencias» de este tipo existen? Los paradigmáticos convencidos de las ciencias sociales no siempre se ponen de acuerdo. ¿Tal vez hay dos: la cualitativa y la cuantitativa, positivista y pospositivista, o realista e idealista? ¿Quizá tres: las teorías realista, hermenéutica y crítica? ¿O posiblemente cuatro: las teorías positivista, pospositivista, constructivista y crítica?

Que no desespere el lector ni decida saltarse este apartado, porque la cuestión concreta afortunadamente es un poco más sencilla. ¿Es que los evaluadores, para ser coherentes, deben decidirse por una u otra? ¿Hay alguna incoherencia inherente en el hecho de aceptar la visión del mundo de uno de los planteamientos, pero emplear los métodos de otro?

Filstead (1979), por ejemplo, defiende una decidida postura paradigmática, como medio para fomentar la evaluación cualitativa. Piensa que las clásicas discusiones entre realismo e idealismo están en la raíz del debate. La fe en la razón vino a expresarse mediante la fe en la ciencia y se ejemplificaba en el positivismo lógico. Para aquellos influidos por el pensamiento idealista europeo, la mente es la fuente del conocimiento y el mundo social lo crean los individuos que viven en él. La evaluación cuantitativa representa el primer enfoque paradigmático, y la evaluación cualitativa, el segundo. Jacob (1987) adopta una postura paradigmática en su tan a menudo citada reseña de las tradiciones de la investigación cualitativa de los estudios sobre educación norteamericanos. Sin embargo, la autora critica que a la investigación cualitativa se la trate como a un enfoque singular, y recurrir a la idea de «tradiciones» de Kuhn para identificar cinco escuelas de pensamiento distintas. Su tesis es que quienes se dedican a la investigación aplicada deben ajustarse a una tradición: «Lo previsible es que los estudios que no sigan una tradición tengan un enfoque

muy pobre, unos conceptos confusos y una deficiente aplicación a la práctica» (Jacob, 1987, pág. 38).

Pensemos un momento en la distinción entre planteamientos cuantitativos y cualitativos de la evaluación de la práctica. Hammersley no está de acuerdo con la postura paradigmática. Sin embargo, resume con sentido práctico las distinciones que otros hacen entre los elementos de cada extremo, unos elementos que se pueden representar en la tabla 3.1. Guba sostiene que todo este tipo de paradigmas responden a tres cuestiones básicas (Guba, 1990, pág. 18):

- ¿Cuál es la naturaleza de la realidad? Una cuestión de *ontología*.
- ¿Cuál es la naturaleza de la relación que se establece entre el conocedor (el indagador) y lo conocido (o conocible)? Una cuestión de *epistemología*.
- ¿Cómo debe proceder el indagador para descubrir el conocimiento? Una cuestión de *metodología*.

TABLA 3.1. *Paradigmas cuantitativo y cualitativo.*

	Cuantitativo	Cualitativo
Los datos son	Cuantitativos	Cualitativos
Trabajo de campo en	Enclaves artificiales	Enclaves naturales
La atención empírica se centra en	La conducta	El significado
El modelo de las ciencias naturales	Se acepta	Se rechaza
El análisis se realiza por	Deducción	Inducción
El proceso teorizador busca	Leyes	Patrones
Filosofía	Realista	Idealista

Fuente: Hammersley (1992, cap. 10).

En la exposición siguiente seguiremos las cuatro preguntas básicas de Guba para distinguir tres posibles posturas paradigmáticas: el realismo pospositivista, la evaluación crítica y la evaluación constructivista. Cada una de ellas supone una diferencia importante en temas como la objetividad (pospositivismo realista), la defensa del papel del evaluador (evaluación crítica) y el relativismo (constructivismo). Pero las tres posturas tienen una coherencia general y se corresponden estrechamente con los principales desarrollos de la filosofía de la ciencia (Bernstein, 1976; Outhwaite, 1987; Phillips, 1987a). Debemos tener en cuenta que el prefijo «pos» se utiliza de forma diferente y poten-

cialmente confusa en la literatura. Por ejemplo, en el término *pospositivismo*, «pos» se usa para expresar que se trata de una posición que mantiene algunas de las premisas ontológicas y epistemológicas del positivismo, con unos cambios más o menos radicales hacia las doctrinas del realismo y la subjetividad. Por consiguiente, «pos» sirve para indicar un mayor o menor elemento de *continuidad* con la posición a la que se refiere. Sin embargo, en el término *posmodernismo*, «pos» se usa para expresar que se ponen en entredicho los móviles y los *motivos* clave de las ideas de la modernidad. Lo que se cuestiona es aquello que une a quienes adoptan esta postura. «Pos», en este contexto, sirve para indicar el elemento de *discontinuidad* con la postura a la que se refiere.

El pospositivismo realista

«Positivista» es un cómodo término despectivo que se ha usado con exageración. Sin embargo, pocas veces atrae adhesiones públicas. Es, como hace mucho que señaló Raymond Williams (1976), «una palabra malsonante con la que nadie que la use emplea un lenguaje soez». A pesar de que hay opiniones contrarias, el positivismo ha ido en declive al menos en los últimos cincuenta años. Los ataques a los profesionales o a los académicos por ser «positivistas» normalmente carecen de fundamento y son perjudiciales (Shaw, 1999). Miles y Huberman admiten que son «positivistas de frente flexible» (1994), con lo que indican sencillamente su compromiso con una versión más rigurosa del realismo. El positivismo ha dejado de ser una elección seria dentro de la investigación de evaluación.

Para el *positivista*, «la tarea de la ciencia es descubrir la verdadera naturaleza de la “realidad” y cómo funciona “realmente”» (Guba, 1990, pág. 19). Tiene también una epistemología objetivista y la idea de un punto de apoyo de Arquímedes. Parece que Arquímedes dijo que con una palanca lo suficientemente larga y un punto donde apoyarla levantaría el mundo: «La objetividad es el punto de apoyo de Arquímedes [...] que permite que el investigador desentrañe los secretos de la naturaleza sin alterarlos en modo alguno» (Guba, 1990, pág. 19).

Phillips (1990a) da una explicación verosímil de las razones para abandonar el positivismo. Tres avances han sido fundamentales para generar entre los científicos sociales la creencia casi unánime en que no existen unas justificaciones absolutas de las afirmaciones científicas.

cas. En primer lugar, se ha revalorado el papel de la observación como último árbitro. Por ejemplo, la aceptación de que algunos mecanismos no se pueden observar llevó a rechazar la creencia de que los conceptos se pueden reducir a una serie de afirmaciones operacionales y observacionales. La misma influencia ha tenido el rechazo del supuesto de que la observación puede ser neutral en lo que a teorías se refiere. El reconocimiento de que toda observación está «cargada de teoría» se comparte en las tres posturas que aquí se describen. Hay mucho más que ver que lo que el ojo percibe. El «justificacionismo» —la idea de que conocimiento equivale a conocimiento demostrado— se rechaza casi de forma unánime. Pero los pospositivistas realistas coinciden en sostener que la observación cargada de teoría no conlleva un relativismo y que es posible mantener una versión de una idea de correspondencia de la verdad: que la investigación puede representar unos fenómenos sociales que son independientes de ella (Hammersley, 1995, cap. 4; Phillips, 1990b).

En segundo lugar, se demostró que la relación entre teoría y observación es más compleja de lo que se pensaba. Quedó claro que las teorías son «poco determinadas» por naturaleza, de forma que nunca se puede decir que se cuenta con la mejor teoría, y se puede construir toda una diversidad de teorías que son igualmente compatibles con las pruebas disponibles. Esto ha conducido al rechazo unánime del «fundamentismo» —la idea de que las conclusiones de la investigación cuya validez sea indiscutible pueden ser el fundamento de la política—. Sin embargo, hay que señalar que esto no ha de llevar a creer que las teorías son siempre falibles, mientras que su base empírica no lo es. Lakatos (1970) llama a esta postura «falsacionismo dogmático». Rechaza con acierto la caricatura de que «la ciencia crece por el repetido derrocamiento de teorías con la ayuda de la dura realidad» (1970, pág. 97). Este escepticismo respecto al valor de la teoría simplemente reinstaura una nueva versión de la observación libre de teorías. Las observaciones son tales a la luz de una teoría dada y «no hay ni puede haber sensaciones que no estén impregnadas de expectativas» (1970, pág. 99). Por consiguiente, aunque no existen justificaciones *absolutas*, esto no significa que *no* existan justificaciones. Pero, si bien podemos tener unas garantías al sostener determinadas ideas, no podemos afirmar que algo sea verdad, que nuestra garantía sea incuestionable, o que siempre existirá: «*Nada* puede garantizar que hayamos alcanzado la verdad» (Phillips, 1990a, pág. 43). Siempre que Popper habla de «experimento», de «base empírica», de «observación», de «teorías aplicables», etc., curiosamente los términos

han de aparecer entre unas comillas invisibles. «Si una teoría es falsada, se demuestra que es falsa; si se “falsa” puede aún ser verdadera» (Lakatos, 1970, pág. 108; Shaw, 1996a, cap. 6). Se han criticado los argumentos de Popper respecto a la falsación. No obstante, la tesis de que la ciencia (y, por consiguiente, la investigación de evaluación) debe ser a la vez crítica y falible es fundamental, como compromiso y como tensión, en todo pospositivismo realista.

En tercer lugar, la idea de que existe una constante acumulación de descubrimientos y de que, de este modo, la ciencia progresa, la han puesto en entredicho de forma efectiva Kuhn, Popper, Lakatos y otros.

La continuidad entre positivismo y pospositivismo no se debe exagerar. No hay duda de que Guba (1990) juega con comentarios irónicos cuando dice que la única alternativa al relativismo es el absolutismo. Phillips insiste con razón: «En ningún sentido [esta] nueva filosofía de la ciencia [...] está íntimamente relacionada con el positivismo» (1990a, pág. 39). Los evaluadores pospositivistas y los influidos por ellos buscan unos equilibrios parciales, del rigor para conseguir una relevancia, de la precisión para conseguir una riqueza, de la elegancia teórica para conseguir una aplicabilidad local, y de las mediciones de resultados para favorecer la investigación del proceso, el significado y el contexto social. Sin embargo, se mantienen distantes tanto de los evaluadores críticos como de los constructivistas (Phillips, 1987b).

La evaluación crítica

Teoría crítica, o más específicamente *evaluación crítica*, es una expresión comodín que incluye la evaluación neomarxista (Anderson, 1989; Lather, 1991; Popkewitz, 1990), algunas posturas feministas (Dullea y Mullender, 1999; Swigonski, 1993), la obra de Paolo Freire y algunas formas de indagación participativa (Reason y Rowan, 1981). Estos enfoques son críticos en el sentido de que los problemas se conceptualizan como parte de los patrones sociales, políticos y culturales en que se forma el evaluando. La investigación crítica se centra en las contradicciones de la práctica. De ahí que la lógica básica no sólo se preocupe de la organización formal de la argumentación, «sino también de las formas concretas de razonamiento que se centran en el escepticismo hacia las instituciones sociales» (Popkewitz, 1990, pág. 49). Por ejemplo, la evaluación crítica de la escuela

no sólo versa sobre el aprendizaje, sino que tiene unas implicaciones que continuamente relacionan la evaluación con cuestiones más generales de producción y reproducción social (Carr y Kemmis, 1986). Esto normalmente supone una «inversión» que conlleva «hacer frágil la historia [...] y así, marcar puntos débiles en la causalidad a la que nos enfrentamos es algo cotidiano, y esto limita nuestras posibilidades» (Popkewitz, 1990, pág. 49).

Por consiguiente, el conocimiento crítico que se consigue con la evaluación nunca es neutral: «Siempre es para algún sujeto particular» (Comstock, 1982, pág. 374). La producción de conocimiento es la producción de valores. Esto tiene unas importantes consecuencias para el papel del evaluador:

Los científicos sociales son partícipes del desarrollo sociohistórico de la acción y la comprensión humanas. Como tales, deben decidir los intereses que van a atender. La única actividad legítima del científico social es participar en la empresa colectiva de ilustración progresiva (Comstock, 1982, pág. 377).

En este aspecto, existen unas diferencias internas dentro de la evaluación crítica. Quizá para la mayoría de los representantes de esta postura; el papel partidista de la ciencia impone la obligación de buscar compromisos políticos mediante la participación activa en los movimientos políticos. Esto «supone una implicación directa y explícita en los esfuerzos por transformar las actuales relaciones sociales» (Popkewitz, 1990, pág. 49). Por citar de nuevo la propia postura de Popkewitz, «los científicos sociales son partidistas en la formación de agendas sociales mediante las prácticas de la ciencia» (Popkewitz, 1990, pág. 50). Así pues, el activismo es implícito, y con él la investigación desvela cómo han de cuestionar las personas el mundo y situarse ellas mismas en las relaciones en curso.

En el ámbito de la investigación y la evaluación críticas existen animados debates. Los marxistas critican con dureza la tendencia *a priori* teorizante de los teóricos críticos y exigen que «los investigadores críticos practiquen en su quehacer empírico lo que predicán en sus formulaciones teóricas» (Lather, 1986b, pág. 258). Otros insisten en que hay que cuestionar tanto la investigación tradicional como la emancipadora, «y se puede descubrir que *ambas* tienen sus carencias» (Humphries, 1997, 4.6). Los evaluadores críticos deben reconocer «la permanente parcialidad del punto de vista de quienes pretendemos construir una investigación emancipadora» (1997, 4.8).

Existen diferencias en el grado en que los evaluadores críticos insisten en el concepto marxista clásico de falsa conciencia. Es probable que quienes ponen mucho énfasis en la falsa conciencia den el mismo peso a las condiciones estructurales, mientras que quienes ponen poco énfasis o ninguno en la falsa conciencia darán una importancia fundamental a los significados intersubjetivos. En efecto, quizá no sea demasiado exagerado decir que la postura que se adopte en este punto es el determinante más importante de la actitud del teórico crítico ante la epistemología y la metodología. En el capítulo 4 hablaremos de nuevo de algunos de estos temas.

El constructivismo

Tal vez como mejor se conozca el tercer paradigma, el *constructivismo*, sea a través de la obra de Guba y Lincoln sobre la evaluación de cuarta generación (*Fourth Generation Evaluation*, 1989). El perfil que en el capítulo 2 hemos dado de su obra servirá para esbozar la postura constructivista/interpretativa. La obra de estos autores ha dado inevitablemente mucho de que hablar, tanto por su postura epistemológica como por su estrategia metodológica. El constructivismo se asocia también con los nombres de Smith y Schwandt. Este último pretende reformular la evaluación en términos de actos prácticos de interpretación, por eso a menudo se describe como un planteamiento interpretativo o hermenéutico. El autor critica el pospositivismo por su presunta modernidad:

Nos oponemos a tres tendencias de la modernidad: 1) [...] concebir el conocimiento sólo como la adquisición del poder para controlarse uno mismo, la sociedad y la naturaleza; 2) [...] presumir que en el método está la clave del conocimiento; 3) [...] definir al conocedor individual como alguien que tiene el control completo de sí mismo, de la naturaleza y de las disposiciones sociales (Schwandt, 1997, pág. 78).

El enfoque que defiende implica

1) que el mundo social [...] sólo se puede estudiar desde una postura de implicación «dentro» de él, y no como un «intruso»; 2) que el conocimiento de este mundo es un conocimiento práctico-moral y, para su eficacia práctica, no depende de la justificación ni de la prueba; 3) que no estamos en una relación de «propiedad» respecto a ese conocimien-

to, sino que lo encarnamos como parte de quienes somos y de lo que somos (1997, pág. 75).

Los defensores de los tres paradigmas aceptan la idea de que la observación está cargada de teoría. Sin embargo, el constructivismo defiende con entusiasmo la carga teórica y de valores de los hechos y la naturaleza interactiva de la indagación, por la que el conocedor y lo conocido «se funden en un todo coherente» (Guba, 1990, pág. 26). Por consiguiente, el interés del constructivista que trabaja desde una postura estrictamente relativista no es tanto comprender con mayor exactitud como entender con mayor profundidad, de forma más verdadera. La investigación, por tanto, no es una tarea epistemológica, sino una actividad moral y práctica. Ningún significado tiene un privilegio epistémico: «La idea que se tenga de la comprensión marca una gran diferencia en la idea que se tenga de la indagación cualitativa» (Smith, 1993, pág. 199).

Aunque existe una interesante diversidad de posturas constructivistas, todas estas citas recogen de forma práctica los *motivos* recurrentes del constructivismo basado en el paradigma. Esta postura se hace oír cada vez más en las publicaciones sobre evaluación, estimulada por la «Conferencia sobre paradigmas alternativos» (Guba, 1990), aunque es posible que estas fuentes tengan menos resonancia en Gran Bretaña que en Estados Unidos.

La tarea del evaluador constructivista consiste básicamente en contar historias:

Las historias son lo más cerca que podemos llegar de la experiencia, cuando nosotros y los demás contamos nuestras experiencias [...] Según esta teoría, las historias que las personas viven son la experiencia. Las personas viven historias, y cuando las cuentan las reafirman, las modifican y crean otras nuevas (Clandinin y Connelly, 1994, pág. 415).

Una gran parte del debate en torno al constructivismo se centra en el esquema relativista, y en la verosimilitud y la coherencia con que se defiende el relativismo (Heap, 1995; McLennan, 1995). También se ha dicho que los constructivistas se comportan regularmente como ambidiestros en su forma de enmarcar el debate. McLennan, por ejemplo, sugiere una crítica extensa de cierto posmodernismo feminista, alegando que las ideas opuestas se tachan permanentemente de «absolutistas». Se lamenta el autor de que se trata de «un trunca-

miento casi escandaloso de los muy diversos esfuerzos filosóficos que han intervenido en este tema [...] y una polémica exagerada hasta el punto de que la erudición simplemente no aparece en la imagen» (McLennan, 1995, pág. 401). En la segunda parte de este capítulo nos extendaremos más sobre este tema.

Volviendo a nuestras preguntas iniciales, ¿es que, para ser coherentes, los evaluadores y los profesionales deben escoger un paradigma? ¿Existe alguna incoherencia inherente al hecho de adoptar la visión del mundo de una teoría y emplear los métodos de otra? Lincoln responde con elocuencia en sentido afirmativo:

La adopción de un paradigma impregna literalmente todos los actos, incluso los que sólo se relacionan de forma tangencial con la investigación, de modo que cualquier consideración, aunque esté remotamente asociada a los procesos de la indagación, exige un nuevo razonamiento para ajustar las decisiones a la visión del mundo que el propio paradigma encarna (1990, pág. 81).

Smith y Heshusius llegan a una conclusión similar cuando responden con un «no» matizado a la pregunta de si los investigadores cuantitativos y cualitativos pueden incluso hablar entre sí: «Al final, es posible que ambas partes acaben casi por hablar distintas lenguas: una lengua científica o libre de valores frente a la lengua cargada de valores del discurso cotidiano» (1986, pág. 11). Los constructivistas no son los únicos defensores de la postura paradigmática. Popkewitz reconoce que el acuerdo se produce en los niveles prácticos de las costumbres de contratación de las universidades y en el debate académico, pero respecto al nivel teórico, dice: «Imagino que aquí es donde queda el pluralismo [...]».

Incluir un sentido disciplinado de la historia en la metodología [...] supone negarse a «ver» los acontecimientos diferenciados, tanto si se adoptan técnicas «cualitativas» como «cuantitativas», sin considerar la relación de los acontecimientos con las formulaciones históricas (Popkewitz, 1990, pág. 65).

El pragmatismo

Sin embargo, se pueden oír réplicas sonoras en sentido negativo desde diferentes posiciones. La postura tradicional de los teóricos de

la evaluación es una mezcla de ideas pospositivistas y pragmáticas (Cook, 1985; Patton, 1990; Reichardt y Cook, 1979; Scriven, 1997; Shadish, 1995). Chelimsky representa esta postura tradicional cuando dice:

Lo que debe contar no es el método que apoye un determinado grupo, sino cómo respondemos a la pregunta o alcanzamos el objetivo [...] poco importa qué perspectiva o qué propósito se haya adoptado: la honestidad es fundamental y la defensa realmente no es una opción (1997, pág. 109).

Algunos etnógrafos defienden también el pragmatismo. Hammersley (1995), por ejemplo, se opone enérgicamente a los argumentos pragmáticos en defensa de la validez de la investigación. Sin embargo, se opone también a las ideas paradigmáticas y sostiene que, para cada uno de los siete elementos de la línea divisoria entre los métodos cualitativos y cuantitativos (véase la tabla 3.1, pág. 74), las características de cada elemento paradigmático siempre se pueden identificar en ejemplos de investigación realizada con el paradigma alternativo. Miles y Huberman advierten también de la tentación presente en los debates epistemológicos de «operar en los extremos»: «En el mundo real de la investigación empírica, creemos que todos nosotros —realistas, interpretativistas y teóricos críticos— estamos más cerca del centro, con múltiples solapamientos» (1994, págs. 4-5). Tal vez el etnógrafo más conocido en este tema sea Howard Becker. Al hablar de las cuestiones epistemológicas, dice: «Creo que es inútil intentar solucionarlas [...] Son, sencillamente, lugares comunes, en el sentido retórico, del discurso científico de las ciencias sociales, el marco en el que se desarrolla el debate. Que así sea [...] Nada hay de trágico en ello» (Becker, 1993, pág. 219). Deberíamos adoptar una perspectiva empírica en este tipo de cuestiones y tratarlas como «un tema, más que como algo fastidioso» (1993, pág. 222). Pero debemos tener cuidado con el efecto paralizante que produce el debate excesivo sobre metodología: «Deberemos seguir trabajando en la teoría, pero no es necesario que pensemos que somos especialmente virtuosos cuando lo hacemos» (1993, pág. 221). Más que considerar que este trabajo teórico es responsabilidad de todos los investigadores cualitativos, el autor prefiere considerarlo una especialidad, la profesión de «la preocupación filosófica y metodológica» (1993, pág. 226).

También en Eisner y Stake se puede encontrar cierto respaldo al pragmatismo. Eisner pregunta «qué puede ser más esclavizante» que

tener «un esquema, un lenguaje [...] un tambor a quien estudiosos y profesores [...] puedan oír», y señala que «es más una cuestión de ver qué es lo que funciona, qué parece adecuado para las situaciones particulares» (Eisner, 1991a, pág. 211). Ya hemos visto en el capítulo 2 que Stake rechaza el uso del término «modelo» porque genera falsas esperanzas.

El pragmatismo resulta atractivo por varias razones y da lugar a parejas inesperadas. Hay que distinguir entre pragmatismo *metodológico* y pragmatismo *filosófico*. El primero se basa en una impaciencia respecto a la filosofía y en un énfasis en la evaluación y la práctica del mundo real, y sostiene que los métodos se pueden separar de la epistemología de la que han surgido. Se insiste, pues, en la utilidad práctica y en la credibilidad de los métodos empleados. La mayor parte de las defensas del pragmatismo a las que nos hemos referido son de este tipo. Por ejemplo, Patton defiende lo que él define como un «paradigma de elecciones»: «Un paradigma de elecciones rechaza la ortodoxia metodológica a favor de la *adecuación metodológica* como criterio básico para juzgar la calidad metodológica» (1990, págs. 38-39).

Gage (1989) vislumbra el momento en que triunfarán las resoluciones pragmáticas de las diferencias de paradigma, como consecuencia de habernos percatado de que las diferencias de paradigma son consecuencia de que los investigadores simplemente estudian temas distintos, a menudo desde una perspectiva disciplinar diferente. Las orientaciones y las reglamentaciones de la educación profesional quizás en todos los países de habla inglesa tienen una amplia base pragmática, a pesar de las presiones de quienes rigen en el mundo de la evaluación. Para todos los que adoptan esta postura, «la metodología concreta empleada es menos importante que el planteamiento general de la acción y la investigación» (Kelly, 1985, pág. 144).

Lo habitual es que el pragmatismo metodológico lleve a rechazar el «purismo» tanto epistemológico como metodológico, a sentirse satisfecho con una metodología «lo bastante buena», y a un enfoque pluralista de los métodos científicos. Así pues, al quejarse de los «filosóficamente obsesos», Scriven dice: «Es mejor edificar sobre lo que posiblemente sea arena [...] que no edificar [...] Es una pérdida de tiempo intentar resolver los problemas de la epistemología sin seguir trabajando» (1997, págs. 478, 479). Reid defiende tesis parecidas, aunque con un tono más moderado, respecto a la asistencia social: «No hay duda de que pueden existir conflictos irreconciliables en el discurso de los filósofos, pero sus perspectivas no son esencia-

les para la tarea de resolver las diferencias y construir un consenso en los mundos reales de la asistencia social» (1994b, pág. 478). Reid quiere «redefinir la naturaleza de las tendencias generales, para que la metodología sea parte, y no se aparte, de ella [...] Ningún método es inherentemente superior a otro, sino que cada uno ofrece al investigador diferentes herramientas de investigación» (1994b, pág. 477).

El pragmatismo metodológico y el *pragmatismo filosófico* tienen unas preocupaciones que se solapan. Sin embargo, existen unas diferencias importantes. El pragmatismo filosófico se basa de forma especial en la obra de los filósofos estadounidenses, y se ha visto reforzado con la obra reciente de Richard Rorty (1979, 1991), cuyas tesis han ejercido una influencia cada vez mayor:

Cuando se encuentra ante una afirmación respecto a un conocimiento, al pragmático le preocupa menos si es verdadera y, en su lugar, se pregunta: «¿Qué supondría creerlo?», «¿qué ocurriría si lo hiciera?», «¿a qué me obligaría?»... Esto cambia el objetivo de la indagación desde la verificación y el recurso al método, hacia la práctica y el recurso a la deliberación y la conversación (Schwandt, 1993, pág. 18).

Reid aplicó al trabajo social una versión menos relativista de algunas de estas ideas. Piensa que tal postura «nos llevará más lejos que las grandes epistemologías que facilitan afirmaciones abstractas sobre la Verdad, la Realidad, etc.», y se sitúa «en la tradición del pragmatismo estadounidense, en especial tal como se establece en la obra de Peirce, James, Dewey y Rorty» (Reid, 1994b, pág. 472). Él y Zettergren (1999) desarrollan la interesante idea de la práctica empírica como una *perspectiva* que permite el intercambio con los enfoques humanista y de defensa.

Las tesis pragmáticas se han empleado también en algunos trabajos del posmodernismo feminista, una de las parejas inesperadas a las que antes me refería. El elemento de sorpresa surge del miedo que el radical tiene de que el pragmatismo tienda a la complacencia política:

Característico de estas metodologías es su orientación a la toma de decisiones y, por consiguiente, a la gestión, su énfasis fundamental en la producción de información útil, su base de valores práctica y pragmática, y su postura metodológica ecléctica (Greene, 1994, págs. 532-533).

EJEMPLO 3.1. ¿PUEDE SER RADICAL EL PRAGMATISMO?

Rorty cree que la filosofía pragmatista puede ser útil para la política feminista. Confía en que el feminismo considere la posibilidad de abandonar el realismo y la idea de que la subordinación de las mujeres es *intrínsecamente* abominable. También le urge a que abandone la pretensión de que existe algo llamado «derecho» o «justicia» o «humanidad» que siempre ha estado de su lado.

Es moderado respecto a la naturaleza de tal empresa: «Uncir el feminismo con el pragmatismo es como uncir el cristianismo con el platonismo [...] algo grande e importante, una inmensa esperanza social, se une con algo comparativamente pequeño y sin importancia» (Rorty, 1991, pág. 6).

Aplicar de este modo el pragmatismo es algo utópico más que radical. Los radicales piensan que hay un error fundamental en las propias raíces:

Los pragmatistas no pueden ser radicales en este sentido, pero pueden ser utópicos. No consideran que la filosofía ofrezca los instrumentos para una cirugía radical [...] La función de la filosofía es, al contrario, la de despejar el camino para los profetas y los poetas, hacer un poco más sencilla la vida intelectual para quienes vislumbran nuevas comunidades (1991, pág. 6).

Por consiguiente, se puede apelar al coraje y a la imaginación: «En mi opinión, la filosofía se hace sentir en otras filosofías, pero no en los problemas sociales como tales» (1991, pág. 12).

La aplicación de estas ideas más cuidadas del pragmatismo a la evaluación está aún en su infancia. Pero tienen el potencial de integrar de forma fundamental en la evaluación aspectos de los enfoques pospositivista, constructivista y de potenciación. Mi postura está en una versión sólida del realismo *fallible* del pospositivismo, el carácter *construido* de la realidad, y el papel básico de los *intereses* políticos y personales. Respecto a los paradigmas, mi premisa es que «la relación entre los paradigmas y los métodos es una cuestión empírica» (Pitman y Maxwell, 1992, pág. 732). Pero creo que Hammersley exagera la debilidad de la relación entre epistemología, valores y métodos. Las acciones de las personas, tanto en la evaluación como en cualquier otra actividad, *están* configuradas por los valores y las visiones del mundo, y las posturas paradigmáticas *no* tienden inevita-

blemente a la intolerancia de la que Atkinson y sus colegas recelaban (Atkinson, Delamont y Hammersley, 1988). Los profesionales y los evaluadores a veces han sido culpables del pragmatismo ingenuo que dice «si funciona, es verdad». Estoy de acuerdo con Greene cuando disiente del pragmatismo metodológico que afirma que la pureza epistemológica no hace que se realice la investigación. Al contrario, «la integridad epistemológica sí consigue que se realice una investigación significativa» (Greene, 1990, pág. 229).

Filstead, uno de los primeros defensores de los paradigmas, conviene en que «lo óptimo es la postura intermedia de mezclar los recursos de ambos enfoques a lo largo de la evaluación», y un defensor tan entusiasta como Guba acepta que existe hoy «una postura más ecuménica» en el reconocimiento compartido de que tanto los métodos de verificación como los de descubrimiento pueden funcionar en todos los paradigmas (Filstead, 1979, pág. 43; Guba, 1990, pág. 23). La tesis de Guba y Lincoln es que el desacuerdo está en los axiomas básicos, y no en las posturas metodológicas. Desde una posición pospositivista, Reichardt y Cook probablemente aciertan al detenerse en el punto en que «los paradigmas no son el único determinante de la elección de los métodos», y cuando concluyen que existe un «vínculo real aunque imperfecto entre paradigma y método» (1979, pág. 16).

Esta tensión esencial entre los paradigmas filosóficos y la práctica es probable que siga siendo objeto de disputas: «Importa a los evaluadores porque produce unos efectos importantes en la forma de entender y realizar nuestro trabajo» (Greene, 1996, pág. 537). Al hablar de los enfoques paradigmáticos hemos mencionado las cuestiones del realismo y del relativismo, del equilibrio entre la relevancia y el rigor, y de los usos de la evaluación. Esta amplia exposición de los paradigmas y del pragmatismo indica el tipo de soluciones que previsiblemente voy a sugerir para cada uno de estos temas.

REALISMO Y RELATIVISMO

Iniciábamos el capítulo con el dilema del realismo y el relativismo tal como se les plantea a los profesionales en su práctica cotidiana. ¿Los profesionales trabajan con unas pruebas bastante directas de la realidad, con constructos realistas, o con los puntos de vista de los poderosos? Como ocurre con la palabra «paradigma», «realismo» es una palabra que ha atraído unos significados cada vez más variados y diferenciados. Al hablar de una variedad de realismo, Leplin señala

irónicamente: «Al igual que el Movimiento por la Igualdad de Derechos, el realismo científico es una postura mayoritaria cuyos defensores están tan divididos que parecen una minoría» (1984, pág. 1).

Del realismo aplicado a la evaluación se suele suponer que exige una idea absolutista del papel de la indagación evaluativa, una idea a la que se denomina fundamentismo. Se cree que éste implica la estrictamente correspondiente idea de verdad, la idea de que la evaluación puede representar al completo las realidades independientes del proceso indagador. Ya hemos señalado que esta forma ingenua de realismo es como un testafierro, que mantiene la adhesión positivista al papel de la observación como reveladora de la verdadera naturaleza de la realidad, y una confianza en la objetividad del evaluador. Sin embargo, esto no descarta la presencia de versiones sólidas del realismo en algunas variedades de la evaluación cualitativa. Blumer y Harré son tal vez los representantes más conocidos. Por ejemplo, Blumer sostiene que una visión constructivista

no traslada la «realidad», como muchos concluyen, del mundo empírico al reino de la imaginación y la concepción [...] El mundo empírico puede «responder» a la imagen que de él hagamos o a lo que de él afirmemos, una respuesta en el sentido de cuestionar y poner en entredicho nuestras imágenes o nuestros conceptos de él, o de no someterse a ellos. Esta resistencia le otorga al mundo empírico un carácter irreductible que es el signo de la realidad (1969, pág. 22).

El constructivismo no es incompatible con el realismo. En efecto, se trata de repetir lo que ya hemos dicho sobre el pospositivismo realista en el contexto de los paradigmas. En esta visión de las cosas podemos seguir creyendo en la existencia de fenómenos independientes de lo que digamos sobre ellos, y en la posibilidad de conocerlos, «sin presumir que podemos saber con certeza si el conocimiento que de ellos tenemos es válido o no» (Hammersley, 1992, pág. 50).

Un realismo crítico o falible es la postura de la mayoría de quienes se identifican con el pospositivismo, además de Blumer y Harré. Es también la postura de la mayoría de los teóricos de la evaluación, como Scriven, Campbell, House, Cook, Shadish y, con cierta ambigüedad, Stake:

La esencia de esta postura es que, aunque existe un mundo real regido por las causas naturales, a los seres humanos les es imposible percibirlo debido a su deficiente mecanismo sensorial e intelectual (Cook y Campbell, 1979, pág. 29).

Dentro de la evaluación realista falible coexiste toda una variedad de opiniones. Por ejemplo, el término «objetividad» significa cosas distintas para diferentes realistas, como ocurre con la idea de «causa». Algunos no se alejan mucho de una posición positivista estricta respecto a su aparente creencia en que la evaluación puede dar una explicación muy aproximada a la realidad. Otros, como este autor, presumen que, aunque cierta versión de la objetividad sigue siendo un «ideal regulador» (Phillips, 1990a, pág. 43), los intereses, la ubicación social del evaluador y de los poderosos interesados y, en expresión sorprendentemente realista de Guba, «la propensión a confundir de la naturaleza», siempre hacen peligrar los procesos y los resultados evaluativos.

Esta modestia epistemológica no se traduce necesariamente en unas aspiraciones modestas respecto al cambio social. Por ejemplo, el carácter tan falible de Campbell se vincula a una audaz idea de una sociedad experimentadora y, desde la asistencia social, Reid pregunta: «¿Es mejor hacer un progreso limitado pero bien documentado, o trabajar con unas metas más importantes con menor certeza de que hayamos conseguido lo que nos proponíamos?» (1988, pág. 45). Pero los evaluadores realistas falibles son menos proclives a asegurar confiados que han conseguido lo que se proponían. Es probable que coincidan con Phillips cuando éste admite que «la objetividad de una investigación no garantiza su verdad [...] *nada* puede garantizar que hemos alcanzado la verdad» (Phillips, 1990a, pág. 43).

Los avances más influyentes de los últimos veinte años han surgido del campo del realismo científico. Los defensores más destacados en el ámbito de la evaluación son House, en Estados Unidos, y Pawson y Tilley, en Gran Bretaña (House, 1991b; Pawson y Tilley, 1997a). El realismo científico es un campo complejo que contiene algo más que su parte de contradicciones internas. Leplin, por ejemplo, hace una relación de seis afirmaciones que son características del realismo, pero no es probable que ningún realista refrende la mayoría de ellas. Sin embargo,

[...] Lo que los realistas sí tienen en común es la convicción de que el cambio científico es, a fin de cuentas, progresivo, y que la ciencia hace posible el conocimiento del mundo más allá de sus manifestaciones accesibles y empíricas (Leplin, 1984, pág. 2).

El contrapunto a este continuo compromiso con una base realista de la evaluación es una epistemología radicalmente relativista de la

evaluación, configurada en torno al supuesto de que la verdad es una cuestión de consenso, que los hechos no tienen significado fuera de cierto esquema de valores y que las causas y los efectos sólo existen en aquello en que se imputan. Los evaluadores críticos rechazan de forma inequívoca la idea de una visión de correspondencia de la verdad. Ésta exige un «acceso independiente a ambos dominios de la mente y una realidad de existencia independiente», dicen Smith y Heshusius, «que no podemos tener» (1986, pág. 10). Los relativistas creen que los problemas, los descubrimientos y las soluciones de un contexto no se pueden generalizar a otro. Los evaluadores son socios subjetivos de los interesados en la creación literal de datos, y actúan como orquestadores de un proceso negociador. Por lo tanto, los datos de la evaluación no tienen ningún estatus especial, sino que simplemente representan otra construcción que hay que tener en cuenta en la marcha hacia el consenso. Así pues, la rendición de cuentas se configura con múltiples elementos, ninguno de los cuales se puede aislar para dar razón de él.

Las principales críticas a la evaluación relativista son similares a las que se hacen al relativismo de las ciencias sociales en general. Tienen el conocido problema de ser autorrefutables, en el sentido de que, si son verdad, son aplicables a sí mismas «[...] y por lo tanto son verdad en relación con una determinada cultura o un esquema particular, y pueden ser falsas desde la perspectiva de otras culturas u otros esquemas» (Hammersley, 1992, pág. 49). Hammersley sostiene también que uno de los efectos del relativismo es que sus partidarios «tienden a oscilar entre la intolerancia indiscriminada y el dogmatismo ideológico». Esto último es un riesgo constante, dicen él y Gomm, porque no existen razones dentro de una comunidad concreta para oponerse a la imposición del paradigma (Hammersley y Gomm, 1997, 3.3). Esto puede llevar a un punto muerto, que a veces se da en la evaluación, en el que cada parte puede alegar que sus opiniones son «verdaderas».

Otro problema concreto relacionado con la evaluación relativista es que las construcciones se alejan de cualquier punto de anclaje en estructuras o intereses. Parecen poco más que constructos personales y mentales, como cuando Guba concluye que «las realidades existen en forma de múltiples construcciones mentales [...] Es la mente lo que hay que transformar, no el mundo real» (1990, pág. 27). Pero la realidad no es simple ni exclusivamente un producto mental. Abarca los actos sociales y colectivos, y estos actos colectivos son transacciones materiales con el mundo.

La posición relativista confiesa de autores como Guba, Lincoln, Smith y Schwandt también plantea unas dificultades especiales para los evaluadores, consecuencia de la presión a que éstos se ven sometidos por tener que buscar para su trabajo unas orientaciones y un respaldo que les den credibilidad y veracidad:

Los evaluadores sienten una particular presión para acogerse a este tipo de procedimientos porque los contextos de la evaluación de programas siguen exigiendo unas garantías de calidad metodológica y de integridad de los datos en el trabajo evaluativo. Es posible que este trabajo no aporte nada a la política social y a la toma de decisión programática, a menos que se perciba como creíble y veraz (Greene, 1994, pág. 537).

Como consecuencia, al relativismo le molestan las luchas del evaluador. Refiriéndose a la evaluación cualitativa, Greene señala:

Muchos profesionales cualitativos luchan con la disonancia invocada por la supuesta dependencia de la mente de toda pretensión de conocimiento social ante las exigencias contextuales (y también personales, relacionadas con el yo) de *«hacerlo bien»*, de *«averiguar qué ocurre realmente en este enclave»* (1996, pág. 280).

La última tesis que se ha desarrollado contra ciertas formas de evaluación constructivista y relativista, y de investigación de los problemas sociales, es que son culpables de lo que se ha denominado «manipulación ontológica» (Woolgar y Pawluch, 1985). El tema exige cierta extensión para poder explicar el argumento exacto implícito. Los argumentos constructivistas (o definicionistas, como prefieren llamarlos Woolgar y Pawluch) respecto a los problemas sociales analizan las definiciones cambiantes de las conductas y las condiciones, desde el punto de vista de las actividades de elaboración de afirmaciones de las personas implicadas. Una determinada conducta o condición desviada —por ejemplo, el abuso sexual de niños, la prostitución, la drogadicción— en la sociedad se define como algo que adquiere un nuevo o un mayor estatus de problema. La tesis definicionista más común es que, dado que la condición básica de fondo no ha cambiado, las variaciones en la definición de la condición deben derivarse de las circunstancias sociales de los definidores, y no de la propia condición: «La separación la marca la presunción de que la propia condición no cambia» (Woolgar y Pawluch, 1985, pág. 215). El problema que plantea esta estrategia explicativa es «cómo se las

arreglan los autores para presentar como objetivas las afirmaciones sobre las condiciones y las conductas, al tiempo que se relativizan las definiciones y lo que en esas afirmaciones se sostiene»:

Una buena explicación de los problemas sociales depende de cuestionar el estatus de verdad de ciertos estados de las cosas seleccionadas para su análisis y explicación, al tiempo que se sitúa en segundo plano o se minimiza la posibilidad de que los mismos problemas sean de aplicación a los supuestos de los que depende el análisis (1985, pág. 216).

Esta estrategia de crear y gestionar las fronteras entre los supuestos que se piensa que son problemáticos y los que no se consideran como tales es lo que Woolgar y Pawluch describen como «manipulación ontológica». Los constructivistas que tienen esta forma de proceder no saben reconocer que sus afirmaciones respecto a la constancia de la condición de fondo deben estar ellas mismas construidas como una afirmación definicional.

Heap describe este problema en términos de monismo y dualismo. El significado del monismo que Heap recoge es la posición metafísica de que sólo existe un tipo de cosas, mientras que el dualismo es la postura metafísica que defiende la existencia de dos cosas (mente/materia, cuerpo/alma, realidad/representación, etc.). Por ejemplo, la posición de la evaluación realista antes descrita en este apartado es claramente dualista: hay conocimiento y hay «conocimiento» (representación). Por otro lado, una posición relativista estricta es monista. Los evaluadores se ocupan tanto del *dominio* de la indagación (el evaluando) como de la propia *indagación*. El relativista coherente «considera que el dominio de la investigación consiste en construcciones, y entiende que la propia investigación consiste en construcciones» (Heap, 1995, pág. 55). No existe una realidad objetiva que sirva de referente para juzgar lo que el investigador afirma.

La opinión de Heap, que queda ilustrada a la perfección en los escritos de Guba y Lincoln, es que existen unas profundas incoherencias entre la retórica y la práctica de la evaluación de cuarta generación. Aunque dicen constantemente que el dominio de la indagación es construido, su metodología se asume siempre en términos dualistas. Ya hemos visto un ejemplo de esto en la frase de Guba respecto a «la propensión a confundir de la naturaleza» (Guba, 1990, pág. 19), donde da por supuesto un mundo que no es simplemente una construcción mental (o social). Heap proporciona algunos ejemplos más

de la presunción de Lincoln y Guba de la existencia de unos evaluandos independientes e intersubjetivamente accesibles en el dominio de la indagación.

Woolgar y Pawluch, y Heap concluyen que la incoherencia que aquí se manifiesta no se puede erradicar: «Quizá toda pretensión de rendir cuentas (explicar) depende de presentar al menos cierto estado de las cosas como objetivo» (Woolgar y Pawluch, 1985, pág. 224). Es posible que el relativismo selectivo del tipo que Heap identifica en Guba y Lincoln no consista en unas «meras dificultades técnicas de las discusiones sobre los problemas sociales, sino [en] unas características omnipresentes de cualquier intento de explicar los fenómenos sociales» (Heap, 1995, pág. 224). Heap se salta el «quizá» de Woolgar y Pawluch:

La evaluación, y la política social, se pueden aceptar de forma coherente y convincente sólo desde una posición construccionista social dualista. [...] Las *investigaciones de evaluación son necesariamente dualistas*. La idea monista de que no existen unos evaluandos reales, sino sólo construcciones de ellos, implica decir que no hay nada que evaluar, a menos que se quieran evaluar las construcciones de los evaluandos (Heap, 1995, pág. 58).

Los argumentos en contra de la evaluación relativista normalmente son convincentes. Sin embargo, también la evaluación realista presenta graves dificultades respecto a sus supuestos sobre el carácter progresivo general del conocimiento científico.

¿RIGOR O RELEVANCIA?

La alternativa de Reid entre objetivos limitados y bien documentados, o de proponerse unos objetivos más importantes con menor certeza, ilustra uno de los «equilibrios» fundamentales al que parece que se enfrentan los evaluadores. Algunos sostienen que la objetividad y la validez son los imperativos superiores del evaluador. Para el evaluador externo, «tanto el distanciamiento como la objetividad siguen siendo unos ideales correctos y muchas veces alcanzables» (Scriven, 1997, pág. 483). Los argumentos de Scriven a favor de mantenerse distante, de realizar una evaluación sin metas, de la prioridad de la evaluación sumativa sobre la formativa, y de rechazar «el llamado diseño participativo» por ser «tan descuidado como se pueda

ser, carente de una autoría participativa del informe final» (1997, pág. 486), no es probable que cuenten con el favor de los evaluadores cualitativos. En mi opinión son o desatinados o todo un ensueño. Pero siguen siendo una tendencia importante en las tesis del rigor evaluativo.

Sin embargo, soplan aires a favor de abordar el tema de la relevancia también, y en algunos casos, en vez de, las ideas tradicionales de rigor. Curiosamente fue la sincera duda de autores como Campbell la que abrió el camino de estos cambios. Por ejemplo, en 1974, cuando «muchos de nuestros alumnos mejores y más trabajadores optan cada vez más por el modelo cualitativo y humanista» (1979, pág. 49), Campbell expuso su insistencia en que el conocimiento cualitativo y de sentido común es el elemento constituyente y la prueba del conocimiento cuantitativo.

En la corriente tradicional de la evaluación se ha abandonado por cierto tiempo la idea de que los diseños experimentales y cuasi-experimentales y los diseños de sistemas singulares pueden producir una información fiable acerca de las causas y los efectos. Se acepta de forma amplia que participamos en «la indagación social por y para los terrícolas» (Cronbach, 1986).

La posibilidad de una evaluación rigurosa también ha estado condicionada por la pérdida de la fe en la racionalidad de la toma de decisiones organizacional, y por el trabajo extensivo realizado por primera vez en los años setenta sobre la naturaleza de la configuración de políticas, y la forma en que la información se usa en esos procesos. De ahí surgieron otras ideas sobre cómo se usa y se puede usar la evaluación para cambiar los programas, los proyectos y las prácticas. El problema de la relevancia se agudiza en la evaluación cualitativa por el enfoque ideográfico de ésta. Para hacer relevantes las conclusiones hay que recurrir a los argumentos adecuados sobre la generalización, o a la inferencia a partir del diseño y el proceso con que se llevó a cabo la evaluación.

Así pues, ¿dónde debemos situarnos en lo que a la relevancia se refiere? La pregunta implica seis aspectos diferentes. Hablamos de los cuatro primeros en las últimas páginas de este capítulo, y en el capítulo 4 lo hacemos del quinto y el sexto.

1. ¿Cómo surgen y se reconocen los problemas de la evaluación?
2. ¿Deben ocuparse los evaluadores de desarrollar una teoría?
3. ¿Qué relación tiene el conocimiento científico con el profesional y de sentido común?

4. ¿Se pueden generalizar los resultados de la evaluación a otras situaciones o prácticas similares e incluso un tanto diferentes?
5. ¿Cuáles son los criterios de una buena evaluación?
6. ¿Cómo se emplean, o se deberían emplear, los resultados de la evaluación?

En primer lugar, la relevancia implica considerar cómo surgen los problemas de la evaluación y cómo se reconocen. Las tesis de la evaluación del profesional y de la investigación en la acción se basan muchas veces en la defensa de una mayor relevancia y en que «toda pérdida de rigor en la investigación queda más que compensada por lo que gana la aceptabilidad de ésta» (Kelly, 1985, pág. 148). El desarrollo de una «evaluación receptiva» por parte de Stake y otros ha ampliado también la implicación de los interesados en los planes prácticos del establecimiento de problemas.

En segundo lugar, el tema del rigor frente a la relevancia se ha planteado en diferentes direcciones por la cuestión no resuelta de si los evaluadores deben ocuparse de desarrollar una teoría. Por teoría entendemos una teoría sustantiva. Greene ya se lamentó antes de que:

Los debates actuales sobre los enfoques cualitativos de la evaluación están llenos de un ruidoso diálogo sobre la teoría metodológica y la práctica, y sólo de algún susurro de interés por cuál sea la naturaleza y la función de la teoría sustantiva o de programas (1993, pág. 26).

Es difícil definir la «teoría», por supuesto. En un sentido importante ya hemos dicho, a partir del hecho de que la observación esté cargada de teoría, que aquí no existe una distinción entre la teoría y el hecho. Y aunque se deje en suspenso esta consideración, las dificultades pueden seguir existiendo porque la teoría se puede referir a proposiciones formales, a una visión del mundo o a una hipótesis de trabajo. «No se pueden recoger piedras en un campo sin contar con una teoría», como parece que dijo William James.

Mi respuesta a la cuestión de si los evaluadores cualitativos deben implicarse en el desarrollo de una teoría sustantiva es «en algunos casos». No acepto la tesis de algunos teóricos de la evaluación de que el desarrollo de la teoría corresponde a la investigación, no a la evaluación. Pero el tema del desarrollo de la teoría plantea unas cuestiones especiales para la evaluación. No sirve que los etnógrafos se limiten a golpearles los nudillos a los evaluadores cualitativos por no ser unos buenos etnógrafos teorizantes (Atkinson y Delamont, 1993). Gree-

ne nos invita a que expliquemos nuestras propias predisposiciones teóricas, describamos las teorías de programa locales (las «perspectivas teóricas significativas sobre la interpretación de los datos»: Greene, 1993, pág. 38), atendamos los nuevos temas teóricos e integremos la teoría en las conclusiones y las recomendaciones de la evaluación. El problema no resuelto (que Greene reconoce) es de qué forma este proceso teorizador debe reflejar unos propósitos evaluativos contestados.

Quienes dicen llevar a cabo una evaluación ateórica en realidad hacen una de estas tres cosas, o una combinación de ellas:

1. Defienden unas teorías tácitamente, sin reflexionar sobre ellas ni declararlas de forma explícita.
2. Las defienden de forma explícita, pero deliberadamente las ocultan al público.
3. Incorporan a su metodología unos conceptos estructurales que corresponden realmente a la teoría, donde quedan escondidos de su vista y de la nuestra (Garrison, 1988, pág. 24).

Más recientemente, Greene detectó «una apertura de la evaluación cualitativa a las posibles aportaciones que ofrecen las formas teóricas de mirar y ver» (Greene, 1996, pág. 282). En modo alguno se trata de avances prematuros.

En tercer lugar, los avances en nuestra forma de considerar la lógica de la evaluación han contribuido también a que se recurra a unos criterios prácticos de relevancia. Se ha producido y aceptado la influencia de las ideas referentes al razonamiento práctico, la lógica informal y de trabajo, y el impacto estimulante de las ideas neoaristotéticas del razonamiento práctico. Todos estos avances ayudan a establecer una distinción entre el rigor de la investigación de carácter general y el razonamiento y la lógica más prácticos y cotidianos, entre los que se incluye también el razonamiento ético. Hay que señalar cuatro preguntas:

1. ¿Cómo deben considerar los evaluadores cualitativos el conocimiento de sentido común (el propio y el de los demás)?
2. ¿Qué idea tienen de la evaluación los profesionales?
3. ¿Qué se quiere decir cuando se habla del carácter práctico de una tarea?
4. ¿La evaluación del profesional se basa en formas de indagación similares a las del investigador evaluativo o en otras distintas?

El conocimiento de sentido común, en especial el que poseen quienes son los receptores de la evaluación, ha tenido su peso durante cierto tiempo (Campbell, 1978; Mayer y Timms, 1970). Smith no hace sino empujar una puerta ya abierta cuando proclama que los interpretativistas piensan que «es hora de prescindir de la antigua idea de que el conocimiento de los investigadores es inevitablemente superior al conocimiento de otras personas, por ejemplo los padres o los profesores, y, por consiguiente, pueden juzgarlo automáticamente» (1992, pág. 100).

Los antropólogos, como Geertz, indican varias razones de por qué hay que tratar el sentido común como un sistema cultural, «un cuerpo de pensamiento relativamente organizado, y no sólo como lo que cualquier persona preparada y en su sano juicio sabe» (Geertz, 1983, pág. 75). Si es un sistema cultural y no una simple aprehensión de hecho de la realidad, entonces «hay en él un orden ingenerado, que se puede descubrir empíricamente y formular conceptualmente» (1983, pág. 92). Aquí es, desde luego, donde el argumento «se hace sentir» en la evaluación cualitativa. La pretensión de tratar seriamente el sentido común debe incluir la pretensión de tratarlo empíricamente. La imagen de sentido común de lo extraño, lo desviado, lo tediosamente prosaico y lo difícil nos da unas razones poco conocidas que, por lo tanto, nos sirven desde una perspectiva antropológica, porque sitúan otras razones cercanas en otro contexto.

Geertz asume esta «disgregación de un concepto a medio analizar» (1983, pág. 93). Dice que si podemos hacer uso del sentido común es porque comprendemos sus «características estilísticas, los signos de actitud que le dan su sello particular» (1983, pág. 85), y no sus variados contenidos. Establece que los atributos generales del sentido común que se encuentra en cualquier parte como una forma cultural son la «naturalidad», la «practicidad», la «ligereza», la «inmetodicidad» y la «accesibilidad».

El sentido común es mucho más que la explicación del ciudadano no instruido ofrecida por el experto profesional o académico. Las recientes reflexiones sobre el lugar de la sabiduría práctica del profesional están indirectamente en deuda con esta idea. Si la evaluación ha de ser relevante, no puede ignorar las formas no instruidas (o instruidas) del gran esfuerzo que los profesionales hacen para comprender desde una perspectiva evaluativa sus actividades prácticas (Shaw y Shaw, 1997a, 1997b). Pero esto plantea otra pregunta. ¿Qué se quiere decir al describir como «prácticas» estas actividades? Al considerar esa pregunta, aparece la renovada influencia de las ideas aristoté-

licas acerca de la teoría y la práctica desde varias instancias, incluidas la etnografía (Hammersley, 1992), el constructivismo (Schwandt, 1993) y la evaluación crítica (Carr y Kemmis, 1986). El principal efecto de esto es doble. En primer lugar, ha conducido a una reinstauración de la dimensión ética del razonamiento y la práctica. En segundo lugar, ha recuperado las ideas de lo práctico de su estatus de segundo escalón, derivativo, y derivado prescriptivamente de la teoría formal. De ahí que Schwab insista en que por práctico «no entiendo [...] lo que se consigue fácilmente, las metas familiares que se pueden alcanzar con medios familiares», sino «una disciplina compleja [...] que tiene que ver con la elección y la acción» (1969, págs. 1, 2).

La reflexión sobre la comprensión de sentido común al final lleva también a la pregunta de si las formas de evaluación del profesional se basan en formas de indagación similares a las que emplea el investigador evaluativo, o en otras distintas. Veremos en el capítulo 6 que la evaluación por parte del profesional plantea toda una diversidad de cuestiones que difieren de los argumentos y las soluciones de cierta teoría de la evaluación de programas anterior.

En cuarto lugar, ¿los resultados de la evaluación se pueden generalizar a otras situaciones o a otras prácticas similares e incluso un tanto distintas? De momento nos limitamos a señalar la importancia de la pregunta formulada. La generalización plantea unos problemas especiales en la investigación cualitativa. Cuanto mayor es el énfasis que se pone en lo local y lo contingente, más cautelosos seremos inevitablemente a la hora de extrapolar las conclusiones a otros lugares y a otros momentos. No hay duda de que existen formas y maneras de abordar la tarea, algunas de las cuales se prestan de forma especialmente adecuada a la práctica evaluativa. Erikson, por ejemplo, propone unas comparaciones estimulantes entre la inferencia realizada en la investigación y la realizada en el ámbito clínico (Erikson, 1959). En un sentido similar, Geertz indica que la tarea de generalizar dentro de los casos, más que entre ellos, es exactamente como el proceso de inferencia clínica (1973, pág. 26). En el capítulo 9 tendremos algo que decir al respecto.

La conexión con las cuestiones de la relevancia frente al rigor se pone de manifiesto en la distinción familiar entre errores de Tipo I y de Tipo II: los falsos positivos y los falsos negativos. Cronbach se lamenta en algún lugar de que los investigadores evaluativos han sido demasiado propensos a aceptar los errores de Tipo II (falsos negativos) por una mala interpretación del riesgo de los errores de Tipo I. Ya hemos señalado en el capítulo 2 que está preparado para equili-

brar la precisión («fidelidad») con la relevancia («amplitud de banda»). Hay muchas más cosas en el cielo y en la tierra de las que soñamos en nuestras hipótesis, diría él, sin duda. Una vez que nos concentramos en conseguir el control mediante la mejora de la validez interna de las evaluaciones, es probable que la relevancia se resienta: «La validez externa», la validez de las inferencias que trascienden de los datos, es el quid de la acción social, no la «validez interna» (Cronbach y otros, 1980, pág. 231).

A los investigadores de evaluación cuantitativa les gusta acusar de falta de precisión a los investigadores cualitativos, por no saber traducir sus afirmaciones a las típicas cantidades. Pero la evaluación cualitativa debe mantenerse en su posición. La acusación no distingue entre precisión y una exactitud más amplia. Lo importante es esta última. Las respuestas pueden ser precisas, pero falsas.

Hemos considerado tres preguntas que están en la base de la comprensión de las pruebas que se obtienen con la evaluación cualitativa. ¿La evaluación se debe basar en un determinado paradigma de práctica evaluativa, o en unas iniciativas pragmáticas e incluso oportunistas? ¿Se puede descubrir una imagen objetiva y real del mundo, o todas las explicaciones evaluativas son relatividades contingentes? ¿Se debe planificar la evaluación de acuerdo con las exigencias del rigor científico o de la relevancia de la práctica? Pasamos ahora a centrarnos en preguntas más «orientadas al uso».

- ¿Cómo se sabe si se hace bien la evaluación?
- ¿Los resultados de la investigación evaluativa se emplean mediante la aplicación directa a los problemas, o mediante un proceso de explicación más general?
- ¿Qué principios éticos y qué reglas de decisión deben guiar la evaluación cualitativa?
- ¿Cómo deben configurar las respuestas que se den a estas preguntas la metodología que se utilice?

CAPÍTULO 4

VALORES, VALIDEZ Y LOS USOS DE LA EVALUACIÓN

Las preocupaciones de que nos ocupábamos en el capítulo 3 son fundamentales en todas las formas de investigación disciplinada de las ciencias sociales. Las particularidades de los paradigmas, el realismo y el relativismo, el rigor de la indagación, las manifestaciones de la relevancia, etc., variarán en función de si lo que interesa es la investigación básica, la investigación aplicada, la investigación de políticas o la evaluación. Pero los argumentos básicos son exactamente los mismos, cualquiera que sea la forma de investigación que se ponga en el microscopio. No se puede decir lo mismo de la mayoría de los temas de que se habla en este capítulo. En cada caso se configuran, se limitan o se enriquecen en virtud del propósito de la evaluación. Las cuestiones relativas a la validez de la indagación, los usos que se le pueden dar a la investigación evaluativa, el enfoque general de los métodos de investigación, e incluso, en cierta medida, las cuestiones de la ética de la evaluación, adquieren un giro distintivo según sea el objetivo de la evaluación. Con esto no se defiende ningún elitismo territorial. No digo que los temas de la evaluación sean más complejos, más importantes, ni siquiera más interesantes que los correspondientes temas de la investigación cualitativa. Pero son *diferentes*.

VALIDEZ Y VALORES

¿Cómo decide y aplica el evaluador los criterios de una buena evaluación? De forma tácita hemos dado por supuestas diferentes respuestas a esta pregunta en el capítulo anterior. Por ejemplo, veíamos que Cronbach, al negociar el equilibrio entre rigor y relevancia, siempre prefiere las respuestas menos fiables sobre una amplia variedad de preguntas: «La calidad científica no es el criterio principal; una evaluación debe fijarse como meta ser comprensible, correcta y completa, y creíble para todas las partes» (Cronbach y otros, 1980, pág. 11).

Campbell, por otro lado, pone el énfasis precisamente en la calidad científica de la evaluación. Su obra sobre la validez interna, y tal vez en especial su explicación de las amenazas a la validez, siguen siendo la caja de resonancia para todo trabajo evaluativo, incluso cuando se modifican o se rechazan sus distinciones originales. Así pues, las elecciones y los equilibrios respecto al rigor y la relevancia suponen unas preguntas sobre cómo se juzga la validez del trabajo de evaluación.

La validez está en la base de la «profesión de preocupación filosófica y metodológica» de Becker (1993, pág. 226), hasta el punto de que cualquier intento de etiquetar las disposiciones de alguien sobre el tema corre el riesgo de arrojar fuego más que luz. Esta disensión en parte es producto de unos parámetros o criterios contestados. Pero el *nivel* en que se emite el juicio también deja lugar para la confusión. Tomemos como ejemplo la evaluación del trabajo en el aula. Cuando preguntamos si esta evaluación es válida, probablemente nos referimos a algo de lo siguiente:

- La validez con que los evaluadores sacan conclusiones de su indagación.
- La validez con que los profesores sacan conclusiones de esos estudios.
- La validez con que los profesores sacan conclusiones de su propia experiencia.

Sin embargo, se puede decir con seguridad que cualquiera que realice una evaluación por escrito adoptará una de las cinco orientaciones siguientes respecto a las cuestiones de la validez:

1. Confianza en el rigor y la validez internos, incluidas las tesis respecto a si la evaluación es fiable, confirmable y se puede generalizar.
2. Confianza en unos criterios más amplios de verosimilitud, credibilidad, validez externa y relevancia.
3. Criterios interpretativos que parten de la idea de que la indagación de evaluación no es una tarea técnica, sino una actividad moral y práctica.
4. La validez juzgada de acuerdo con el cumplimiento de un modelo de evaluación como servicio o reforma.
5. Una evaluación abiertamente ideológica, en la que ésta se juzga de acuerdo con su efecto catalizador sobre la vida de los que

no tienen poder alguno, y por el cumplimiento de una función de defensa por parte del evaluador.

Cuando se dice que se adoptará *una* de estas posturas, hay que dejar claro que se refiere a la argumentación o la retórica que se adoptan, y no necesariamente a la posición práctica aplicada a una determinada evaluación. Los evaluadores pueden cambiar de una posición a otra, o contar con elementos de más de una posición al mismo tiempo. La probabilidad de que así ocurra es mayor cuando dos o más investigadores de evaluación trabajan en equipo. Sin embargo, se pueden hacer unas distinciones generales. La orientación de la validez interna se asocia de forma especial con los diseños experimentales y cuasi-experimentales. El esbozo de la posición de Campbell que hacíamos en el capítulo 2 es suficiente para nuestros propósitos; no pretendo detallarla de forma más minuciosa. Sin embargo, la calidad de «caja de resonancia» de su obra, a que antes nos referíamos, se pondrá de manifiesto más de una vez en los párrafos que siguen.

Verosimilitud y credibilidad

Los evaluadores etnográficos de tendencia general, los evaluadores pospositivistas y aquellos comprometidos con posiciones filosóficas más pragmáticas sobre la evaluación se suelen caracterizar por la *segunda* posición. Hammersley ofrece una de las defensas más completas de esta visión general de la validez.

EJEMPLO 4.1. LA IDEA DE VALIDEZ DE HAMMERSLEY

La postura de Hammersley se ha desarrollado en parte contra argumentos más recientes a favor de la validez basados en epistemologías relativistas y posturas de defensa dentro de la investigación. Sus tesis están recogidas en sus libros *What's Wrong With Ethnography?* (1992) y *The Politics of Social Research* (1995).

La idea general de Hammersley respecto al objetivo de la evaluación es: «La función de la investigación es ofrecer una información que sea a la vez verdadera y relevante para algún asunto público legítimo» (1992, pág. 68). La verdad y la relevancia no siempre están en armonía. Es posible que afirmaciones falsas sean provechosas (útiles),

y no se puede prever siempre la futura utilidad de unas afirmaciones verdaderas pero aparentemente triviales. Pero Hammersley cree que los criterios básicos siguen siendo buenos.

Piensa que las disputas sobre la validez de la investigación se deben resolver desde el punto de vista de su verosimilitud («la coherencia con los conocimientos existentes cuya validez se considera fuera de toda duda razonable»; 1995, pág. 75) y de la credibilidad («la probabilidad de que el proceso que produjo la afirmación esté libre de errores graves»; 1995, pág. 75). La idea de «duda razonable» es importante, en el sentido de que actúa de freno de la posibilidad de una regresión infinita de afirmaciones y contraafirmaciones.

Las cuestiones fundamentales son: ¿la afirmación de la investigación/evaluación es razonable dados los conocimientos actuales? De no ser así, ¿es creíble? Reconoce el autor, en una advertencia significativa, que «aunque [la coherencia con la creencia existente] no descarta la aceptación de nuevas creencias [...] la hace menos probable que en el modelo fundacionalista» (1992, pág. 72). Señala asimismo que se debe dar por supuesta la creencia en que las personas están dispuestas a que se les convenza. La existencia de públicos con diferentes intereses implicará distintos criterios para poder determinar la verosimilitud de las pruebas.

¿La afirmación es fundamental en la argumentación general que se presenta? De ser así, entonces las pruebas tendrán que ser más convincentes. El tipo de afirmación afectará a los criterios empleados. Hammersley distingue entre definiciones, descripciones, explicaciones y teorías. La descripción es local y está vinculada al tiempo. La teoría, ni lo uno ni lo otro:

La valoración de la validez implica identificar las principales afirmaciones que hace el estudio, señalar el tipo de afirmación que representan, y luego comparar las pruebas que se aportan para cada una de ellas con lo que se juzga que es necesario, dadas la verosimilitud y la credibilidad de la afirmación (1992, pág. 72).

Según esto, es claro que, aunque el criterio de validez se considera constante en todos los públicos, su aplicación a casos concretos puede variar considerablemente. Las ideas de Hammersley sobre el proceso por el que se deben resolver las disputas sobre la investigación están configuradas por su forma de entender el discurso racional. Propone la creación de un discurso que tenga las siguientes directrices (1995, cap. 4):

- El interés primordial debe ser la verdad de las afirmaciones, y no sus consecuencias políticas o prácticas.

- Se debe juzgar los argumentos tan sólo por su verosimilitud y credibilidad, y no por las características personales de quienes los propongan.
- Los evaluadores deben estar dispuestos a cambiar sus ideas, y deben comportarse como si los otros investigadores también estuvieran dispuestos a hacerlo, al menos hasta que aparezcan unas sólidas pruebas en contra.
- Cuando no se llegue a un acuerdo, el investigador debe aceptar la existencia de una duda razonable sobre la argumentación que se ofrece.
- No se debe restringir la participación en este tipo de discurso por razones de actitud política o religiosa.

Hammersley elabora deliberadamente estas directrices para excluir las concepciones relativistas y fuertemente activistas del propósito de la investigación. También pretende conservar el estatus de afirmaciones de verdad científicas frente a afirmaciones profanas.

Algunos aspectos de esta argumentación se encuentran a veces como parte de posturas evaluativas un tanto diferentes. Ocurre así porque los investigadores evaluativos que aceptan las formas de evaluación relativistas o las abiertamente ideológicas muchas veces se basan en criterios de validez que corren paralelos a los *principios* convencionales (Lather, 1986a; Lincoln y Guba, 1986a; Reason y Rowan, 1981). Reason y Rowan, por ejemplo, se refieren a las pruebas de capacidad para describir y diferenciar lo que existe como criterio de la validez del conocimiento experiencial, y lo formulan en sentido directamente paralelo al criterio tradicional de validez de la medición. La obra anterior de Lincoln y Guba ofrece una versión cuidadosamente desarrollada de este análogo naturalista de las versiones convencionales del rigor (tabla 4.1, véase pág. sig.). Los criterios de la columna de la derecha se deben leer como «análogos u homólogos metafóricos» de los criterios positivistas (Lincoln y Guba, 1986a, pág. 76). En la *credibilidad* incluyen la implicación prolongada, la observación persistente, la triangulación de diferentes tipos y el control de los miembros. La *transferibilidad* se hace verosímil mediante las pruebas de la «descripción gruesa» (Geertz, 1973) —la narrativa sobre el contexto—. El control externo del proceso de evaluación permitirá al evaluador deducir la *confiabilidad*, y el control externo del producto facilitará la deducción de la *confirmabilidad*.

TABLA 4.1. *Análogo naturalista de los criterios de validez convencionales.*

Valores	Criterios convencionales	Problemas consiguientes	Logrados mediante	Análogos de la veracidad
Verdad	Validez interna	Confusión	Control, aleatoriedad	Credibilidad
Aplicabilidad	Validez externa; generalización	Atipicidad	Muestras de probabilidad	Transferibilidad
Coherencia	Fiabilidad; reproducibilidad	Inestabilidad	Reproducción	Confiabilidad
Neutralidad	Objetividad	Parcialidad	Aislamiento del investigador	Confirmabilidad

La evaluación como actividad moral y práctica

Aunque Lincoln y Guba consideran que sus análogos de la veracidad suponen un avance significativo sobre los criterios de validez convencionales, se han orientado hacia criterios de validez de características más relativistas, anclados en la categoría de *autenticidad*, que en gran parte aborda aspectos de las cuestiones éticas e ideológicas. Este tercer planteamiento del modo de establecer los criterios de una buena evaluación renuncia más o menos a los criterios que significativamente se pueden describir como validez. Se basa en el rechazo del realismo, en el que ningún significado goza de privilegio epistémico, y la «validez» de la evaluación no es una cuestión que tenga que ver con comprender de forma más precisa, sino con comprender de forma más profunda y verdadera. Las pruebas de una buena evaluación se formulan en términos de justicia, de una mayor conciencia de la complejidad, una mejor comprensión de los valores de los demás y un mayor respeto por ellos.

Smith es un defensor inflexible de esta postura. Entiende el ideal regulador de este intento de comprender como una solidaridad humana, y no como una objetividad. Las decisiones procedimentales que uno toma al afirmar la calidad de un estudio son unas decisiones de carácter moral:

Por ejemplo, uno puede asumir el control de los miembros porque es lo que, dadas las circunstancias, corresponde hacer desde un punto de vista moral y ético, y no porque ese control se vaya a traducir en un estudio más objetivo (Smith, 1992, pág. 103).

Por tanto, «nuestros juicios de bondad son logros prácticos, asumidos en el contexto del diálogo y la persuasión, que elaboramos mientras avanzamos» (1992, pág. 103). Como lo expresan Reason y Rowand, «el conocimiento válido es una cuestión de relación [...] más personal e interpersonal que metodológica» (Reason y Rowan, 1981, págs. 241, 244). El enfoque consensual de los juicios de validez significa que, en tal planteamiento, «válido es una etiqueta que se aplica a una interpretación o una descripción en la que uno está de acuerdo» (Smith y Heshusius, 1986, pág. 9) y, por consiguiente, es un término honorífico.

Esta posición no es equivalente a la postura emancipadora, ni siquiera a la reformista. Así, cuando Smith ilustra lo que entiende por solidaridad humana, explica que leer una buena investigación «me ha permitido comprender a las personas y las situaciones que se estudiaron de forma diferente y, espero, con mayor profundidad y sensibilidad» (Smith, 1992, pág. 105). En el capítulo 3 reseñábamos los principales problemas de las posturas relativistas del tipo que Smith adopta. En el contexto actual parece que nos hemos encontrado con otro problema de la relación entre la retórica y la práctica en la evaluación constructivista. La comprensión indirecta y la sensibilidad parecen poca cosa ante toda la retórica ética e ideológica que hay detrás de ellas. El elefante de la retórica ha dado a luz un ratón de la práctica; un ratón deseable pero, al fin y al cabo, un ratón.

La crítica que Woolgar y Pawluch hacen del constructivismo se basa en la incoherencia, en que ese tipo de análisis normalmente presume cierto conocimiento de las condiciones sociales reales. Y suele hacerse con unas afirmaciones relativizadoras, al tiempo que se presume que la situación subyacente permanece inmutable. Después de esta acusación de «manipulación ontológica» se han clarificado varias posturas dentro del constructivismo. Como resultado de ello, Best (1989) sugiere de forma convincente que de hecho existen tres tipos de construccionistas.

En primer lugar, los *construccionistas estrictos*, que aceptan las críticas de Woolgar y Pawluch (1985; y de Heap, 1995; véase también el capítulo 3 de este libro) y, en consecuencia, rechazan hacer ningún tipo de suposición. En segundo lugar, los *construccionistas demoleedores*. Éstos son básicamente objetivistas, en el sentido de que presumen que se produce la construcción social, pero es equivocada. En otras palabras, las pretensiones constructivistas son una distorsión del conocimiento. Entre estas dos posturas probablemente se encuen-

tra, en tercer lugar, la mayor parte de las investigaciones y evaluaciones construccionistas, a las que Best denomina *construccionismo contextual*. Estos constructivistas aceptan que se hacen ciertas suposiciones sobre las condiciones sociales y también que estas suposiciones *se pueden* hacer: «Los construccionistas contextuales sostienen que cualquier afirmación se puede evaluar [...] presumen que pueden conocer, con una confianza razonable, las condiciones sociales» (Best, 1989, pág. 247). Es evidente que en esta posición resuena el realismo falible y crítico de algunos pospositivistas, del que hablamos en el capítulo 3.

La evaluación reformista

La evaluación como servicio es quizá una forma de distinguir la evaluación de la investigación, desde el punto de vista de su *propósito*, más que desde el de su *metodología*. Este cuarto enfoque general de la evaluación y la correspondiente postura sobre la validez las resume Simon como «un servicio práctico, particularista, político y persuasivo» (1987, pág. 8). Después de describir el desarrollo de la evaluación democrática, la autora confiesa que «personalmente, el atractivo del modelo democrático para mí está en su lógica educativa, más que en su política de oposición» (1987, pág. 53). En Gran Bretaña, la postura de investigación de políticas de Finch (1986) refleja el rechazo de la idea de que el evaluador es un técnico, y en Estados Unidos Greene insiste en que «ya no podemos ocultar nuestra condición de ciudadanos tras nuestras subjetividades científicas. Debemos convertirnos en ciudadanos científicos» (1996, pág. 287).

Quienes defienden esta posición muchas veces razonan que la evaluación debe ser educativa y democratizadora. En este planteamiento se pueden encontrar diferentes grados de énfasis en el papel de la defensa, desde el pluralismo de Cronbach y Stake, a las posturas reformistas de House, Simon y Finch. House ha sido durante mucho tiempo el defensor de un criterio de justicia social para la evaluación, basado en un concepto de justicia como honradez e imparcialidad. En otros trabajos más recientes se pretende desarrollar las ideas de justicia como criterio para juzgar la evaluación (por ejemplo, Altheide y Johnson, 1997; Kirkhart, 1995). Aunque existen cuestiones espinosas no resueltas respecto a las diferentes ideas de justicia social, es un campo que recompensa la atención que se le preste. En el capítu-

lo 7 consideramos las cuestiones de la justicia en relación con las decisiones referentes al diseño de la indagación.

Los criterios abiertamente ideológicos

El compromiso con la justicia forma parte también de la quinta postura sobre la evaluación y la validez: la asociada con una evaluación abiertamente ideológica. Sus defensores comparten con los constructivistas la determinación de situar las bases del debate respecto a la validez lejos de la naturaleza del mundo y de cómo podemos conocerlo y colocarlas en las cuestiones de los valores y los principios, donde las referencias morales del evaluador son fundamentales.

¿Qué cabida tiene la defensa mediante la evaluación, esta «polémica de primer orden», como lo describe Chelimsky (1997, pág. 109)? ¿Tiene razón la autora cuando concluye que «la defensa no es realmente una opción» (ibíd.)? ¿Son justos y están en lo cierto Hammersley y Gomm cuando proclaman inflexiblemente que la defensa socava la confianza pública en la imparcialidad de la evaluación y la investigación sociales y,

en la medida en que tales desarrollos equivalen a redefinir el objetivo de la indagación como el fomento de cierta causa práctica o política, nos parecen unas parcialidades motivadas, y creemos que los investigadores deben oponerse a ellas? (Hammersley y Gomm, 1997, 5.4).

Si la evaluación ha de alcanzar la aspiración de Chelimsky de unos criterios de calidad y de práctica que sean «integradores de todos los fines legítimos de la evaluación y que no sirvan meramente para santificar una perspectiva a expensas de otra» (1997, pág. 108), entonces la clase dirigente de la evaluación debe afrontar los temas difíciles, en vez de definir las fronteras para excluirlos.

Lather propone un análisis especialmente creativo de los temas de la validez en la evaluación y la investigación abiertamente ideológicas. Confiesa la autora que «lo que he visto una y otra vez en la literatura metodológica de la investigación abiertamente basada en valores es una falta de claridad acerca de la necesidad de controlar la credibilidad de los datos» (Lather, 1986a, pág. 77). Concluye que los criterios mínimos que la evaluación necesita para incorporar unos diseños con los que abordar la validez son:

- La triangulación de los *métodos*, las *fuentes de datos* y las *teorías*.
- Una subjetividad reflexiva (cierta documentación sobre cómo la lógica de los datos ha afectado a los supuestos del investigador).
- Una validez nominal (que se establece mediante el reciclaje de categorías, la aparición de nuevos análisis y conclusiones a través de al menos una submuestra de encuestados).
- Una validez catalizadora (cierta documentación de que el proceso de investigación ha llevado a unas ideas y, ojalá, a la acción por parte de los encuestados).

Lather advierte de que «si no conseguimos desarrollar estos procedimientos, no conseguiremos proteger nuestro trabajo de nuestras pasiones, y nuestra construcción de una teoría se va a resentir» (1986a, pág. 77). Aborda la autora otro problema difícil: la falsa conciencia. Lo habitual es que los neomarxistas consideren que el evaluador es el intérprete del mundo y quien expone la falsa conciencia. Lather critica las tendencias al elitismo y la alienación que tal idea del investigador genera:

Esta percepción no dialéctica del papel del investigador frustra la intención de desmitificar el mundo para los desposeídos [...] Existe al menos cierta revisión necesaria de la tendencia a rechazar la resistencia a las interpretaciones marxistas como «falsa conciencia» [...] pero las ideas teóricas y empíricas siguen estando destinadas a otros intelectuales (1986a, pág. 75).

En sus propios estudios sobre la resistencia de los estudiantes a un currículo liberador, Lather declinó rechazarla como falsa conciencia. Al contrario: «Quiero analizar qué nos pueden enseñar estas resistencias acerca de nuestra propia tendencia a la imposición» (1991, pág. 76). Reconoce la tensión con la que han de vivir esos investigadores orientados a la praxis:

El desarrollo de una teoría social emancipadora exige una postura empírica incondicional, dialógicamente recíproca, basada en el respeto por la capacidad humana, pero a la vez profundamente escéptica ante las apariencias y el «sentido común» (Lather, 1986b, pág. 269).

Cuestiones sobre la validez

Queda aún una serie de temas difíciles sobre la validez. Aquí nos limitamos a señalarlos. En primer lugar, la cuestión de si los evaluadores han de ser partidistas no va a desaparecer. En los párrafos anteriores he dicho de forma implícita que existen cuatro amplias respuestas a esta cuestión: el no partidismo, el multipartidismo, las posiciones reformistas y los planteamientos abiertamente ideológicos y radicales (véase la tabla 4.2).

TABLA 4.2. *¿Debe ser partidista la investigación de evaluación y políticas?*

Solución	Asociada con
No partidismo	Scriven, Hammersley
Multipartidista	Cronbach, Stake
Reformista	House, Finch
Partidismo radical	Lather, Oliver

En segundo lugar, decíamos en el capítulo anterior que la evaluación cualitativa ha prestado poca atención a la cuestión de si los productos de la evaluación *se generalizan* a otras situaciones. La crítica que Cronbach hace de los argumentos de Campbell a favor de la validez interna ha gozado de un amplio reconocimiento, pese a que aún se conoce muy poco y poco se ha reflexionado sobre ella. Hemos señalado la tesis de Cronbach de que la validez externa —la validez de las inferencias que trascienden de los datos— es el *quid* de la acción social, no la validez interna. Las consecuencias de su posición son metodológicamente radicales. El autor está convencido de que, en última instancia, la formulación de la validez externa de Campbell es trivial y de que la cuestión de la generalización está en cómo generalizar a personas, intervenciones y situaciones *que no formaron parte de la muestra y son diferentes*.

Las epistemologías relativistas, además de una proporción sustancial de los enfoques etnográficos ortodoxos de la evaluación, rechazan las ideas referentes a una generalización basada en inferencias de probabilidad implícitas. Otra proporción menor cree también que la relevancia local de la evaluación es todo lo que tenemos. En el capítulo 9 sostenemos que los evaluadores cualitativos no necesitan, ni se pueden permitir, asumir esta última postura.

En tercer lugar, ¿deben los evaluadores reciclar de forma regular la evaluación con quienes facilitaron la información? A esto se le suele llamar la validación de los miembros. Hay tres posibilidades básicas, aunque no son del todo mutuamente excluyentes. Primera, se puede optar por una interpretación realista de la validación de los miembros (y también de la triangulación) como método para validar los datos. Las ideas del investigador y del encuestado permiten localizar y situar lo que en realidad ocurre. Segunda, se ha dicho que esta idea corroboradora de la validación de los miembros es demasiado ingenua y que las respuestas de los miembros se deben «considerar como otros datos importantes, una ocasión para ampliar y perfeccionar el análisis del investigador» (Bloor, 1997b, pág. 48). Tercera, podemos decidir mantener un proceso de retroalimentación de los miembros y responder no a partir de una base epistemológica, sino de unos compromisos éticos y morales con la evaluación participativa y democratizadora. Esta tesis es similar a la que veíamos en la versión de la validez de Smith. La cuestión se agudiza cuando se planifica el análisis de la evaluación, y la consideraremos con mayor detalle en el capítulo 9.

En cuarto lugar, ¿qué posturas de valor deberían adoptar los evaluadores en lo que a las conclusiones de su indagación se refiere? La línea divisoria en este tema se encuentra entre aquellos que adoptan una posición descriptiva y aquellos que adoptan una posición prescriptiva. Lo habitual es que en una posición *descriptiva* se concluya una evaluación diciendo: «Éstas son las posiciones de valor que adoptan los interesados en esta evaluación, junto con las posibles consecuencias que parecen derivarse si se adopta esta o esa otra postura». Con frecuencia, los evaluadores constructivistas son evaluadores descriptivos, tal y como ocurre en el caso de Stake, para quien lo importante son las decisiones de los interesados. Lincoln y Guba, por ejemplo, afirman que «todas las ideologías deberían tener las mismas oportunidades de expresión en el proceso de negociación de las recomendaciones» (1986a, pág. 79). En una posición de valor *prescriptiva* se busca persuadir a las personas clave de las virtudes o cualquier otra cosa del evaluando, y se hacen recomendaciones en relación con la acción futura. No existe ninguna forma definitiva y rápida de hacer coincidir las posiciones de valor con las posturas sobre el partidismo de la tabla 4.2. Scriven, por poner un ejemplo, es un valorador prescriptivo, como también lo es House. La mayoría de los evaluadores que tienen una posición de partidismo radical también asumen posturas prescriptivas.

Ambos extremos de este debate tienen sus problemas, por lo que se han intentado desarrollar posiciones intermedias. Eraut es un ejemplo de esta posición cuando habla de los temas del valor de la evaluación. Se opone radicalmente a la insistencia de Scriven en que los evaluadores deben recoger su tesis de evaluación en un juicio conciso: «Los valores en conflicto no se pueden condensar sin imponer cierto esquema ético superordenado» (Eraut, 1984, pág. 31). Tampoco acepta una postura de evaluación neutral y descriptiva. La selección y la interpretación de los datos están íntimamente conectadas, de manera que «las explicaciones supuestamente descriptivas encarnan unos principios ocultos de selección. El evaluador salva su propia conciencia a expensas de no ayudar a sus clientes» (1984, pág. 31).

La solución provisional de Eraut es lo que llama una «evaluación divergente»:

Una de las tareas del evaluador es estudiar las conexiones entre las pruebas empíricas y los diversos valores que puedan proporcionar criterios para juzgarlas [...] Al mostrar los vínculos que existen entre las pruebas empíricas y ciertos valores, el evaluador ayuda a quienes defienden tales valores (1984, pág. 32).

Hasta aquí es una postura similar a una versión de la posición descriptiva que se propone en alguna otra parte de la literatura (Shadish y otros, 1990, pág. 101). Sin embargo, Eraut aporta un importante elemento nuevo al añadir que se ofrece un mismo servicio a valores rivales que tal vez no puedan estar representados dentro de la evaluación. Esto permite que las personas puedan poner en cuestión los valores y

que el evaluador introduzca unos valores no considerados por ninguno de los encuestados sin darles ninguna prioridad especial. De modo que el objetivo de la evaluación divergente es aumentar la sensibilidad hacia otros puntos de vista, representar todas las posiciones de valor de las partes implicadas e introducir unos valores adicionales donde corresponda (Eraut, 1984, pág. 32).

La postura de Eraut encaja perfectamente en la posición de multipartidismo de la tabla 4.2 (véase pág. 109).

¿INGENIERÍA O ILUSTRACIÓN?

En el fondo, lo que guía la práctica de la evaluación es cierta visión de finalidad. En el centro de esta visión está la idea de que la evaluación tiene una utilidad para las políticas, los servicios y la práctica. Hasta los años setenta, se presumía que ese uso era relativamente directo y consistía en aplicar los resultados de la evaluación, de una forma directa instrumental o de ingeniería social, al problema que la evaluación abordaba. Esta idea era respaldada en Estados Unidos por los primeros trabajos de Campbell y Scriven y, en Gran Bretaña, estaba representada por el modelo fabiano de la relación entre el investigador y el responsable de formular políticas.

El cambio en los supuestos referentes al equilibrio entre rigor y relevancia esbozado en el capítulo anterior supuso una nueva forma de considerar los usos de la evaluación. Una figura clave en estos cambios fue Carol Weiss. En el perfil que de sus ideas hemos trazado en el capítulo 2, señalábamos su tesis de que las consideraciones políticas se inmiscuyen en la evaluación de tres formas:

1. El programa y las políticas que se evalúan son ellos mismos criaturas de decisiones políticas.
2. La evaluación incluye consideraciones políticas porque sus informes entran en el foro político.
3. Las políticas —normalmente las de la clase dirigente— están implícitas en la investigación de evaluación.

El trabajo empírico de Weiss condujo a teorías más realistas sobre el uso de la evaluación, en contraste con las aspiraciones más utópicas de la idea de Campbell de una sociedad experimentadora. Weiss introdujo las ideas de «acrecentamiento de la decisión» y de «reducción del conocimiento» para explicar que las decisiones políticas no se toman en un solo punto del tiempo, sino mediante la suma de pequeñas elecciones, la reducción de pequeñas opciones y la disminución gradual de las alternativas disponibles (Weiss, 1980). Sostenía que las decisiones de quienes formulan las políticas reflejan un modelo de ilustración del uso, según el cual existe una infiltración difusa e indirecta de las ideas de la investigación en la interpretación que esas personas hacen del mundo. Las ideas del uso como ilustración se han convertido en la ortodoxia de prestigio más reconocido.

La idea que se adopte sobre los usos de la evaluación va a afectar, a su vez, a la forma de entender la tarea de difusión de la informa-

ción. La obra de Weiss y otros autores implica que no debemos considerarla una cuestión de comunicación técnica, sino que va a plantear cuestiones sobre la forma de entender el cambio social. Las profesiones y sus investigadores han tendido a no plantearse problema alguno en su forma de entender el tema de la difusión y de las destrezas para la utilización de la investigación (Fisher, 1997). En el capítulo 5 analizamos las implicaciones que la obra de Weiss, Cronbach y otros tiene para el uso de la información en la investigación y la evaluación de políticas.

Todo esto está relacionado con el tema de los paradigmas del que nos ocupábamos en el capítulo anterior. Es evidente que la postura que se adopte en lo referente al uso de la evaluación estará configurada por la idea que se tenga de ésta. Los realistas pospositivistas son más proclives a adoptar una visión de ingeniería social del uso de la evaluación; los constructivistas quizá consideren el uso como una forma de narración de historias; y los evaluadores críticos entenderán el uso de la evaluación desde la perspectiva de la implicación y la acción políticas (Greene, 1990).

No hay que exagerar el acento sobre la ilustración, como tampoco hay que presumir que las ideas sobre el uso que se han desarrollado en los campos de la evaluación de programas y de investigación de políticas se van a trasladar fácilmente a la evaluación del profesional.

Es más probable que los modelos de *ilustración* estén relacionados con:

- El cambio a largo plazo mediante la investigación de evaluación.
- La evaluación cuyo objetivo es cambiar la forma en que las personas conceptualizan su experiencia.
- La evaluación de programas.
- El teorizar sobre los objetivos de la evaluación.
- La evaluación basada en la enseñanza superior.

Es más probable que las ideas *instrumentales* de la evaluación se asocien con:

- Aplicaciones inmediatas y a corto plazo.
- La evaluación del profesional.
- La teorización limitada sobre ciencias sociales.
- La evaluación local basada en la organización y que pueden realizar «los de dentro».

Al problema del uso se le puede dar un énfasis exagerado. Chelimsky señala con razón que «el uso es algo bueno [...] pero poner en él un celo excesivo nos convierte en personas insulsas e inocuas, cuando en realidad deberíamos cuestionar las políticas dominantes y desafiar el *statu quo*» (1997, págs. 105, 108).

LA ÉTICA DE LA EVALUACIÓN

La ingenuidad respecto a la ética es, en sí misma, inmoral. Si Lincoln y Guba están en lo cierto, existen razones firmes para preocuparse:

Los valores de los encuestados *se desprecian* sistemáticamente como simples opiniones sin base alguna en el conocimiento científico. Los propósitos de la investigación *se ocultan* sistemáticamente a los sujetos alegando que, si los conocieran, quedaría en entredicho la «adecuación técnica» del estudio [...] *Se accede* a información personal protegida cuando el investigador considera que es útil para su búsqueda de la verdad (1989, pág. 225).

Las preocupaciones aquí planteadas se refieren al proceso cotidiano de la toma de decisiones ética, más que a un principio ético más amplio. Pero incluso en un ámbito más general existen diferentes ideas sobre qué principios éticos amplios deben regir la evaluación. Eisner (1991a), por ejemplo, se sitúa en un territorio familiar cuando dice que los tres principios más importantes son el consentimiento informado, la confidencialidad y el derecho a abandonar la investigación en cualquier momento. House tiene una idea un tanto distinta. Señala que los tres principios éticos son el respeto mutuo, la no coerción y no manipulación, y el respeto y defensa de unos valores democráticos (House, 1993). El autor desarrolla estos principios, por ejemplo, con el establecimiento de un acuerdo de evaluación justo (véase el capítulo 7), pero admite que se puede discutir cuáles pueden ser los valores reales que su tercer principio conlleva.

Warwick piensa que los principios éticos se pueden subsumir en el principio único de conservar la libertad humana. Analiza un estudio quizá muy discutido por sus implicaciones éticas: *Tearoom Trade*, de Laud Humphreys. Éste adoptó el papel encubierto de «loca» para estudiar la (por entonces ilegal) actividad homosexual masculina. Observó unos doscientos actos sexuales. Algunos hombres conocían la investigación que estaba realizando, y otros no. Luego obtuvo ilegal-

mente las direcciones de los hombres del estudio y, sin que le reconocieran y alegando que estaba realizando una «encuesta sobre salud», les entrevistó (Humphreys, 1975).

Bajo el epígrafe general de libertad humana, Warwick señala los costes y los beneficios potenciales que para la realización de una investigación puede tener tal libertad. Entre los *beneficios* posibles están la oportunidad de expresarse uno mismo a través de las conclusiones del estudio, el gusto de compartir cosas importantes con otra u otras personas, la satisfacción intelectual de participar en la investigación y los posibles beneficios personales derivados del hecho de hablar del problema. Los *costes* potenciales incluyen el engaño, la invasión de la intimidad y el mal uso de la información. Se pregunta el autor si es necesario cuestionar los supuestos liberales referentes al valor de la información. Pregunta: «¿La investigación refuerza la tendencia de los individuos a ser precavidos y a simular la conducta que de ellos se espera?» (Warwick, 1982, pág. 52). Señala que la investigación consume recursos de dinero, personal, buena voluntad y tolerancia, todos ellos finitos. Aunque es verdad que existe la obligación de estudiar los temas polémicos, esto no incluye el derecho a engañar, explotar o manipular.

Lo que Warwick dice se aplica por igual a la ética de la evaluación. Sin embargo, Guba y Lincoln adoptan un enfoque diferente de la ética y vinculan firmemente su análisis a su posición paradigmática.

EJEMPLO 4.2. LA ÉTICA: ¿HA FRACASADO LA CIENCIA POSITIVISTA?

Lincoln y Guba creen que los fracasos de la toma de decisiones ética apuntan a tres lecciones. En primer lugar, «mucho depende del “punto de ebullición moral” del investigador en cuestión; diferentes investigadores tomarán decisiones distintas incluso ante circunstancias similares» (1989, pág. 226). En segundo lugar, nada que sea inherente a los modelos convencionales de investigación en las ciencias sociales obliga a una conducta ética ni la recompensa. Por último, «los investigadores han conseguido encontrar muchas razones aparentemente sólidas para evitar la “ética de la sabiduría” —las prácticas éticas ideales— en la realización de sus investigaciones» (1989, pág. 226).

La fuerza de su argumentación apunta principalmente al segundo problema, lo que ellos describen como «acometida» ética del paradigma convencional. En esencia, su tesis dice que el paradigma convencional adopta una visión absolutista de la verdad y supone que se pueden conseguir unos conocimientos irrefutables:

Con esta justificación metafísica de la búsqueda de la verdad en la mano, el científico social es libre de defender de forma convincente que su investigación exige y justifica el engaño [...] Los supuestos sobre la naturaleza de la realidad refuerzan, y sin duda exigen, que se trate a los sujetos humanos como si de objetos se tratara (1989, págs. 226-227).

Lincoln y Guba creen que el paradigma naturalista, aunque no esté libre de sus propios dilemas éticos, es más ético que el convencional. Como no tiene una premisa subyacente sobre cómo son realmente las cosas, el engaño es «absolutamente contraproducente» (1989, pág. 230). Además, la relación entre el investigador y el encuestado es una relación entre socios que tienen una misma voz. Por consiguiente, el entrevistado conserva el control del proceso.

La argumentación ayuda a recuperar los temas éticos como algo fundamental en la investigación y en la evaluación. Sin embargo, su análisis plantea varias preguntas:

- ¿Está justificada la postura de Lincoln y Guba en el sentido de que unos paradigmas son más éticos que otros?
- Decidir si los problemas éticos del naturalismo son preferibles a los de la ética de la investigación convencional es una cuestión de juicio. Por ejemplo, Lincoln y Guba dicen que en la indagación naturalista «evidentemente no se pueden garantizar la confidencialidad y el anonimato» (1989, pág. 233).
- Lincoln y Guba coinciden en su crítica del positivismo y del pospositivismo realista. Sin embargo, adoptan posturas muy distintas en lo que a la posibilidad y la naturaleza de la objetividad se refiere. Es posible que su crítica pierda parte de su fuerza una vez que se reconozca esta distinción.
- ¿Es evidente que la indagación naturalista garantiza que no va a haber engaño? Tiene una cierta verosimilitud retórica, pero algunos comentaristas se preguntan si la etnografía puede ser más propensa al engaño que los métodos de indagación tradicionales (Stacey, 1988). Aunque las relaciones igualitarias sí contrarrestan el engaño, un problema, del que hablábamos en el capítulo anterior, es que en muchas investigaciones etnográficas e interpretativas se distingue entre los constructos estables del investigador y los constructos inestables y contingentes del encuestado. Heap indica que tal situación es sospechosa desde un punto de vista moral, pues «a los encuestados se les tratará como a locos metafísicos y dualistas estúpidos, mientras que, por otra parte, el evaluador honesto y franco conoce la verdad: todo son construcciones; nada es real» (1995, pág. 59).

Con esta justificación metafísica de la búsqueda de la verdad en la mano, el científico social es libre de defender de forma convincente que su investigación exige y justifica el engaño [...] Los supuestos sobre la naturaleza de la realidad refuerzan, y sin duda exigen, que se trate a los sujetos humanos como si de objetos se tratara (1989, págs. 226-227).

Lincoln y Guba creen que el paradigma naturalista, aunque no esté libre de sus propios dilemas éticos, es más ético que el convencional. Como no tiene una premisa subyacente sobre cómo son realmente las cosas, el engaño es «absolutamente contraproducente» (1989, pág. 230). Además, la relación entre el investigador y el encuestado es una relación entre socios que tienen una misma voz. Por consiguiente, el entrevistado conserva el control del proceso.

La argumentación ayuda a recuperar los temas éticos como algo fundamental en la investigación y en la evaluación. Sin embargo, su análisis plantea varias preguntas:

- ¿Está justificada la postura de Lincoln y Guba en el sentido de que unos paradigmas son más éticos que otros?
- Decidir si los problemas éticos del naturalismo son preferibles a los de la ética de la investigación convencional es una cuestión de juicio. Por ejemplo, Lincoln y Guba dicen que en la indagación naturalista «evidentemente no se pueden garantizar la confidencialidad y el anonimato» (1989, pág. 233).
- Lincoln y Guba coinciden en su crítica del positivismo y del pospositivismo realista. Sin embargo, adoptan posturas muy distintas en lo que a la posibilidad y la naturaleza de la objetividad se refiere. Es posible que su crítica pierda parte de su fuerza una vez que se reconozca esta distinción.
- ¿Es evidente que la indagación naturalista garantiza que no va a haber engaño? Tiene una cierta verosimilitud retórica, pero algunos comentaristas se preguntan si la etnografía puede ser más propensa al engaño que los métodos de indagación tradicionales (Stacey, 1988). Aunque las relaciones igualitarias sí contrarrestan el engaño, un problema, del que hablábamos en el capítulo anterior, es que en muchas investigaciones etnográficas e interpretativas se distingue entre los constructos estables del investigador y los constructos inestables y contingentes del encuestado. Heap indica que tal situación es sospechosa desde un punto de vista moral, pues «a los encuestados se les tratará como a locos metafísicos y dualistas estúpidos, mientras que, por otra parte, el evaluador honesto y franco conoce la verdad: todo son construcciones; nada es real» (1995, pág. 59).

En nuestras explicaciones y observaciones sobre la ética de la evaluación hemos adoptado deliberadamente una perspectiva general. Los problemas y los dilemas más concretos se consideran en capítulos posteriores, según nuestra idea de que el análisis del propósito de la evaluación debe ser fundamental. Las cuestiones específicas de la ética se contemplan en relación con el diseño de la evaluación, el trabajo de campo y el análisis, en los capítulos 7-9.

METODOLOGÍA Y MÉTODOS

En la teoría de la evaluación existe una ambigüedad en lo que se refiere a las actitudes ante las metodologías y los métodos, que supone a la vez una bendición y una pesadilla. Gran parte de las obras anteriores a los años setenta ha servido para recalcar que la evaluación no está ni debe estar regida exclusivamente por la metodología. La idea de que «el evaluador que sólo esté equipado con la metodología no está bien preparado» (Shadish y otros, 1990, pág. 183) ha entrado a formar parte de la ortodoxia, aunque su aceptación no sea universal. Se recela del purismo metodológico (Hammersley, 1993b).

Hay varias razones para dudar de los métodos:

- En primer lugar, las dudas tienen su origen en parte en el énfasis que, especialmente dentro de la evaluación cualitativa y constructivista, se pone en el uso del yo como instrumento y en la desconfianza en los métodos como una búsqueda moderna de la clave del conocimiento. Los interpretativistas recelan de los investigadores que están «demasiado enamorados de los procedimientos o la metodología» (Smith, 1992, pág. 104). Eisner coincide con esta postura cuando, con cierto desdén, señala que «en cuestiones cualitativas, los libros de recetas no garantizan nada» (1991a, pág. 169). Lincoln y Guba dicen que «el relativismo del naturalismo sugiere que es imposible (y siempre lo será) especificar una metodología definitivamente verdadera para llegar a conocer» (1989, pág. 231).
- En segundo lugar, la duda metodológica puede ser una consecuencia de aceptar que la observación está cargada de teoría. Garrison infiere que sólo la teoría puede convertir los fenómenos en significado. Los métodos «sólo proporcionan la sintaxis, nunca la semántica» (1988, pág. 25).

- En tercer lugar, del cambio hacia una idea ilustrada de la evaluación, según la cual los procedimientos instrumentales ocupan un segundo lugar frente a los temas políticos más amplios y los equilibrios de la evaluación, puede derivarse una menor preocupación por la contundencia metodológica.
- Por último, no se puede ignorar la probabilidad de que la indiferencia metodológica se deba en parte a una simple carencia de imaginación metodológica, muchas veces manifestada en un pragmatismo metodológico ingenuo.

Por estas razones, las implicaciones metodológicas de los temas de que hemos hablado son más generales. Por ejemplo, las grandes líneas metodológicas surgen de las diferentes posiciones paradigmáticas. Este vínculo es completamente manifiesto tanto en la evaluación crítica (Comstock, 1982) como en el constructivismo (Smith y Heshusius, 1986). Deriva también del cambio pospositivista producido en la evaluación de tendencia general. La consecuencia ha sido un énfasis en los métodos múltiples (el «multiplismo crítico», para emplear la fea expresión de Cook; Cook, 1985). Existe un peligro en la insistencia actual en el enfoque del «caballo según carrera» de las decisiones referentes al diseño de la evaluación. Se dice, para presentarlo con cierto grado de caricatura, que los cuasi-experimentos seguirán siendo adecuados cuando se necesite medir las causas, los efectos y los impactos, mientras que los métodos de descubrimiento y los estudios de casos serán los adecuados cuando a los evaluadores se les pida que presenten una imagen a gran escala y que narren historias. Es una situación demasiado cómoda, que se traduce en una clausura excesivamente temprana del debate sobre los métodos y la metodología, en la que ha influido sin duda el cansancio ante lo que en Estados Unidos se ha llamado la «guerra de paradigmas». En efecto, la tesis fundamental de esta postura dista mucho de ser directa. Por ejemplo, la propia obra de Campbell sobre el estudio de casos «construye una de las argumentaciones más sólidas hasta la fecha para demostrar que los métodos cualitativos pueden producir un conocimiento causal válido» (Shadish y otros, 1990, pág. 135). Las cuestiones causales relativas al diseño y el análisis de la evaluación cualitativa son un ejemplo importante de que el propósito de la evaluación configura de forma distintiva la evaluación cualitativa, de forma que tiene menos probabilidades de manifestarse en la investigación cualitativa de carácter tradicional. Así se refleja en el espacio que se dedica a estos temas en los capítulos 7 y 9.

La obra de Cronbach también tiene consecuencias metodológicas. Le lleva a la evaluación *dentro* de los programas, más que a la evaluación *entre* programas. «Su idea de las realidades del programa social le arrastra hacia unas concepciones metodológicamente menos estrictas de los diseños de la evaluación», y «no quiere que ninguna concepción concreta del método científico trivialice el proceso de formular preguntas importantes» (Shadish y otros, 1990, pág. 349). Entiende el diseño de la evaluación como una serie de «equilibrios ingeniosos» (Cronbach y otros, 1980, pág. 225), que no sólo son técnicos, sino que forman parte del acuerdo político. «El *poder de acción* es la base. El poder de acción se refiere a la probabilidad de que la información, *si se cree*, cambie el curso de los acontecimientos» (1980, pág. 265; cursiva del original). Para él esto no significa retroceder a una idea instrumental de cómo se va a usar la evaluación, porque «las cuestiones que mayor poder dan a la acción son las que contribuyen a la reflexión» (1980, pág. 266). Los constructivistas aceptan compromisos similares con unos diseños menos estrictos.

Hemos señalado que la forma de entender los usos de la evaluación, su lógica general, las posturas sobre el paradigma y las ideas sobre la validez y la generalización darán un valor especial a la metodología. Veámos en el capítulo 3 que también el hecho de otorgar a la *teoría* un lugar central en la evaluación nos aleja de la idea de los evaluadores como expertos técnicos. Los evaluadores y los profesionales deberán ser cada vez más conscientes de su postura ante este tipo de cuestiones. Sólo entonces tendrán la posibilidad de resolver las preguntas metodológicas gemelas de la evaluación, es decir, con qué métodos es más probable que se alcancen los propósitos de una determinada evaluación y qué valor especial y qué «ingresos» proporcionará el propósito evaluativo a esos métodos.

En esta segunda pregunta está oculto el problema más general de la relación entre la práctica de la evaluación/investigación y la práctica profesional. Con frecuencia se ha dicho (Bloom, 1999; Gould, 1999) que los dos tipos de práctica son muy afines. Y lo habitual es que se diga que «la enseñanza/la práctica legal/la asistencia social/el diagnóstico médico/la gestión es como la evaluación/la investigación», más que al revés. En otras palabras, parece una forma de cambiar las percepciones que la comunidad tiene de la tarea *profesional*, más que de la tarea de *investigación/evaluación*. Cuenta con una larga tradición. El problema de este tipo de afirmaciones es que se han empleado para apoyar una serie de pretensiones aparentemente incompatibles a favor de los planteamientos de gestión mediante la eficacia, la

práctica científica, la práctica reflexiva, etc. Hace más de veinte años, y en un contexto de trabajo social, Thomas (1978) advertía de la ingenua presunción de un propósito o un método comunes entre investigadores y profesionales. Pero la idea es interesante, y los investigadores y los profesionales deberían seguir explorando las correspondencias y las diferencias entre la investigación y la práctica.

La posición sobre los métodos de evaluación que se adopta en este libro se basa en la idea de que no existen unos métodos de evaluación distintivos. Disiento de la aparente idea de Lincoln y Guba (1986b) de que las técnicas empleadas para llegar a los datos en la investigación, la evaluación y el análisis de políticas raramente se solapan. De esto no se deduce que los evaluadores tengan poco que decir sobre los métodos. En los capítulos 7-9 analizamos el carácter distintivo que los propósitos evaluativos otorgan al diseño, la recogida de datos y el análisis.

CONCLUSIÓN

Las cuestiones generales que hemos considerado en los dos últimos capítulos no están del todo resueltas. En efecto, es probable que Becker esté en lo cierto cuando implícitamente afirma que siguen siendo una costumbre y una práctica dentro de la indagación en las ciencias sociales. Si es así, tal vez convenga que nos autoasignemos una postura prescriptiva. Mi enfoque preferido incluye unos compromisos con:

- El trabajo empírico sobre cómo los miembros de profesiones afines (y de profesiones superficialmente muy diferentes; Schon, 1983) comprenden la práctica a través de la evaluación.
- Un nuevo examen de los riesgos y los beneficios que conllevan las opciones referentes al paradigma y las diferentes formas de pragmatismo.
- Una cautela tanto respecto a los supuestos relativistas como a los muy realistas acerca del conocimiento que proporciona la evaluación.
- Un alejamiento de los juicios globales sobre si la evaluación debe poner el énfasis en el rigor o en la relevancia.
- La dilucidación de los criterios de validez que se aceptan en los proyectos de evaluación cualitativa.
- Un acuerdo previo sobre los procedimientos por los que se pueden alcanzar unos juicios de validez finales.

- La identificación de una serie de expectativas referentes al uso final de la investigación, a las que previsiblemente se van a enfrentar los evaluadores.
- El desarrollo de estudios de casos de los principios y la toma de decisiones éticos.
- La revisión de la adecuación de los métodos de evaluación cualitativos.

Las fuentes empleadas en este libro indican los nuevos caminos más prometedores. La revisión selectiva de algunos de los principales teóricos de la evaluación compensará el esfuerzo. En lo que a mí respecta, creo que Cronbach, Weiss, House, Greene y algunos aspectos de la obra de Lincoln y Guba y de Stake constituyen un terreno especialmente fértil. El diálogo, a ser posible cara a cara, entre los evaluadores que trabajen en diferentes campos profesionales fomentará de forma especial la buena evaluación dentro de los grupos profesionales. Para ello se necesitará la colaboración entre las diversas disciplinas más allá de las fronteras nacionales. El comentario de George Bernard Shaw, tan familiar para los británicos, de que «los norteamericanos y los británicos son dos pueblos separados por un mismo idioma» es una idea aplicable a la semejanza y, también, a las diferencias en cómo se ha desarrollado la evaluación en un país y en el otro.

EVALUAR PROGRAMAS Y POLÍTICAS

La evaluación se puede aplicar a políticas, programas, proyectos o servicios concretos de algún programa o proyecto. Y, en cada caso, se puede estar en diferentes fases del proceso, desde la fase de planificación, pasando por la de viabilidad y ejemplificación, hasta la de la plena realización y financiación. La evaluación no se limita a la medición de los resultados de las políticas y los programas oficiales. Tampoco es su objetivo pretender establecer que un determinado cambio observado es el resultado de una intervención dada. Mi punto de partida en este libro era la afirmación de que entre los propósitos de la evaluación estaban también los de llegar a comprender los asuntos públicos y los intentos actuales y anteriores por abordarlos, el fomento de la práctica reflexiva y potenciadora, y el desarrollo de la cultura de la evaluación.

Sin embargo, de todo esto no hay que deducir que los temas evaluativos sean prácticamente los mismos tanto si se trata de políticas, de programas o de profesionales individuales. Cronbach y sus colegas exageraban muy poco cuando decían que «si existe algún pecado individual responsable del caos actual, es la disposición a hacer afirmaciones generales que supuestamente se aplican a todas las evaluaciones» (Cronbach y otros, 1980, pág. 51).

Pensemos, por ejemplo, en el tema de los usos de la evaluación. En los capítulos anteriores hemos visto que los trabajos de diversas personas, como Weiss y Cronbach, han llevado a la aceptación casi universal de las limitaciones de un modelo de uso basado en la presunción de una aplicación directa e instrumental de las conclusiones de la evaluación a un problema concreto y predeterminado. Pero si se piensa un poco se verá que, en el caso de la evaluación del trabajo de un profesional individual, tendrá poco sentido un modelo de uso de ilustración y a largo plazo, un modelo que podría poner en peligro los derechos al servicio de los pacientes, los clientes o los alumnos en cuestión. Los modelos de uso de ilustración o explicación se han desarro-

llado principalmente en el ámbito de la política y en programas amplios, más que en proyectos y prácticas locales.

En este capítulo y en el siguiente consideraremos algunas de las cuestiones más concretas que se plantean en la evaluación cualitativa aplicada a programas, políticas y prácticas. La evaluación de programas, junto con la metodología de las pruebas aleatorias controladas, se suele considerar aún a menudo como el «criterio de oro [...] el Rolls Royce de los planteamientos de la evaluación» (Chelimsky, 1997, pág. 101). En la primera mitad de este capítulo reflexionaremos sobre algunos de los temas especiales que se nos presentan cuando nos embarcamos en estrategias cualitativas de evaluación de *programas*. El término «programa» se utiliza en un sentido muy flexible en este contexto. Se entenderá que los organismos, los centros educativos o los proyectos teóricamente independientes y no coordinados, que presten unos servicios más o menos análogos o que se enfrenten a una serie de problemas comparables, constituyen un programa. En la segunda parte del capítulo nos fijaremos de forma más general en los debates y las cuestiones que se plantean en la evaluación cualitativa de *políticas*. En el capítulo 6 describiremos y consideraremos algunos de los enfoques más recientes e innovadores de la evaluación cualitativa dentro de la *práctica* directa.

LA EVALUACIÓN DE PROGRAMAS

Joseph Wholey presenta la siguiente definición al principio de su importante manual sobre metodología y métodos de la evaluación de programas:

La evaluación de programas incluye la medición del rendimiento del programa —los gastos en recursos, las actividades del programa y los resultados del programa— y la comprobación de los supuestos causales que vinculan a estos tres elementos (Wholey, 1994, pág. 15).

Lamentablemente, esta definición plantea más problemas de los que resuelve. Quizá su única ventaja es que traza el mapa de lo que es un campo minado. Tiene cuatro problemas evidentes. Primero, se relaciona de forma demasiado explícita con la identificación de los resultados. Se mencionan las actividades del programa, pero la implicación es que la identificación de las relaciones causales *desvelará* la importancia real de estas actividades en relación con los resulta-

dos del programa. Segundo, Wholey subraya acto seguido la contribución de la evaluación de programas a una mejora del rendimiento del programa: «Una aportación potencial importante de la evaluación de programas es el uso que de ella hacen los responsables de las políticas, los gestores y el personal, para cambiar los recursos, las actividades o los objetivos del programa con el fin de mejorar su rendimiento» (1994, pág. 15). Sin embargo, ésta no es sino una de las múltiples razones para llevar a cabo la evaluación. En efecto, Cronbach piensa que «la evaluación de un determinado proyecto tiene importantes implicaciones para los proyectos que se pondrán en práctica en el futuro» (Cronbach y otros, 1980, pág. 267). Los valores de la política se limitan también si se reduce su relevancia a un programa determinado. Tercero, el énfasis fundamental que en su definición se pone en la medición y la ausencia de referencias claras al contexto político más amplio de los programas indica una idea del propósito y el proceso de la evaluación muy simplista. Por último, en la interpretación que Wholey hace de la evaluación de programas hay unas referencias a la gestión que para nada sirven.

Las cuestiones de la evaluación de programas se diferencian mejor si se habla de los diferentes *géneros* de evaluación de programas. Por ejemplo, Greene previó de forma muy aproximada las distinciones que hacíamos en el capítulo 3 de este libro y señaló que son cuatro. Las denomina pospositivismo, pragmatismo, interpretativismo y evaluación crítica o normativa (Greene, 1994). Las preguntas de evaluación típicas que se plantea la evaluación pospositivista son: ¿se consiguen los resultados deseados, y son atribuibles al programa?, ¿este programa es la opción más eficaz?¹ Las evaluaciones pragmáticas se centran en preguntas del tipo: ¿qué partes del programa funcionaron bien, y cuáles se deben mejorar?, ¿qué efectividad tiene el programa respecto a los objetivos de la institución y respecto a las necesidades de sus beneficiarios? Lo habitual es que la evaluación interpretativa empiece por preguntar qué experiencia tienen del programa los interesados. La evaluación crítica y normativa se plantea preguntas del tipo: ¿de qué forma las premisas, los objetivos o las actividades del programa sirven para mantener las desigualdades de po-

1. Greene es una de las autoras actuales más incisivas sobre temas de evaluación. Sin embargo, aquí y en otras partes (Greene y McClintock, 1991), tiende a adoptar una definición del positivismo que parece estar demasiado influida por el énfasis engañoso que Guba pone en la continuidad entre el positivismo y el pospositivismo (Guba, 1990).

der y de recursos en la sociedad?, ¿cómo puede cuestionar estas desigualdades estructurales el proceso de la evaluación?

EJEMPLO 5.1. LA EVALUACIÓN DE PROGRAMAS PRENATALES

En la literatura sobre evaluación se encuentran dos ejemplos que suponen un contraste ilustrativo que se produce de forma natural y que nos puede ser útil. Wholey describe la valoración preliminar de la evaluabilidad del Programa Prenatal de Tennessee. Whitmore también describe la evaluación de un programa educativo para mujeres embarazadas solteras en una zona deprimida de Halifax, en Nueva Escocia (Whitmore, 1990, 1994; Wholey, 1994).

El Departamento de Salud Pública de Tennessee desarrolló un proyecto de cinco años: «Mejora del resultado del embarazo» (TIOP en sus siglas inglesas), desde 1977 a 1982. En su último año, se financió una evaluación del programa. Wholey dirigió una valoración destinada a planificar las actividades de evaluación que pudieran resultar más útiles. Formó grupos de trabajo y de política a los que incorporó a personas encargadas del presupuesto, a gestores y ejecutivos de experiencia, un comisionado de Sanidad, personal experto en proyectos y otras personas. A continuación se propusieron diversas opciones de diseño, entre las cuales estaba:

- Conseguir cumplir los criterios generales del Estado.
- Comparaciones internas del programa entre los proyectos TIOP y los diferentes condados.
- Comparaciones previas y posteriores y análisis puntuales, utilizando los datos referentes a los nacimientos.
- Comparación de datos previos, posteriores y puntuales entre los condados con proyectos TIOP y condados que no contaran con ellos.
- Análisis previos, posteriores y puntuales en que se compararan los proyectos TIOP que ofrecieran diferentes servicios con distintos costes.

A la luz de todo esto, se llegó al acuerdo de que el objetivo debía ser reducir el número de alumbramientos de niños con poco peso, como indicador indirecto del riesgo de morbosidad y mortalidad infantil, y de la discapacidad para el aprendizaje. Los representantes del programa pensaban que era un objetivo a la vez realista y del que podían rendir cuentas. El objetivo de la mortalidad infantil se mantuvo también por indicación del Comisionado de Sanidad, que pensaba que posteriormente se podrían formular preguntas sobre este indicador.

El programa en el que participó Whitmore se llevaba aplicando tres años cuando se propuso su evaluación. Whitmore ofreció facilitar una indagación participativa. Se contactó con todas las mujeres que se pudo que previamente hubieran utilizado el servicio y se las invitó a que solicitaran participar en el equipo de evaluación. Se nombró a cuatro de ellas, con el encargo de formular el diseño, recoger y analizar los datos, y presentar los resultados. Dice Whitmore que la investigación y la evaluación participativas combinaban

la investigación, la educación y la acción, en un esfuerzo por mejorar las condiciones sociales y económicas en las que viven las personas. Sus principales objetivos implican un cambio, una justicia y una igualdad sociales amplios, que se consiguen con la potenciación de los miembros menos poderosos (1990, pág. 216).

La recogida de datos se basaba en entrevistas no estructuradas y cuestionarios. Hubo dificultades comprensibles. A las mujeres les resultaba difícil separar los datos de sus propias opiniones. Las tensiones dentro del grupo eran habituales, y provocaron el abandono de uno de sus miembros. Los contactos con los medios de comunicación demostraron ser «extremadamente destructivos para nuestro proceso colectivo y, en última instancia, para la cohesión del grupo» (1990, pág. 221). También se plantearon problemas de comunicación y confianza. Una mujer dijo:

Nuestro mundo es diferente del vuestro. El miedo se debe a que las personas siempre quieren humillarte, rebajarte [por el hecho de estar en la asistencia social] [...] Nuestro modo de vida es distinto del vuestro. Simplemente, no confiamos en las personas como vosotros confiáis (Whitmore, 1994, pág. 92).

Pero los resultados globales fueron positivos. Tuvieron contactos provechosos con el Grupo Asesor, dieron una conferencia y asistieron a una clase en la universidad. Además, Whitmore y el grupo concluyeron que se habían conseguido unos beneficios individuales, para el grupo y para el entorno social más amplio (véase la tabla 5.1).

TABLA 5.1. *Resultados de la evaluación del programa prenatal.*

Nivel de práctica	Elemento potenciador
Individual	autoconfianza conocimientos/destrezas entusiasmo y placer

Grupo	superación del aislamiento desarrollo de la confianza
Entorno	oportunidades educativas hablar con grupos externos cuestionamiento del sistema de servicios sociales oportunidades de empleo

Estas breves descripciones de cada proyecto subrayan el hecho de que las evaluaciones de programas que parten de esquemas distintos varían en los valores claves que fomentan, el público al que se dirigen, los métodos preferidos y las preguntas de evaluación típicas que se formulan.

TEMAS PARA LOS EVALUADORES CUALITATIVOS DE PROGRAMAS

Aunque no existe un buen modelo de práctica exclusivo para la evaluación cualitativa de programas, hay una serie de preguntas cuyas respuestas establecen los parámetros generales de una postura cualitativa sobre la evaluación. Los evaluadores cualitativos de programas:

- Considerarán el *proceso* de las actividades y de la aplicación del programa antes de considerar sus resultados.
- Desconfiarán de la aplicación y la evaluación del programa, que dependen en gran medida de los supuestos de la *toma de decisiones racional*.
- No deben ignorar la importancia de *teorizar* sobre los programas.
- Serán menos entusiastas del valor de los modelos de evaluación de *rendición de cuentas*.
- Incluirán diversas posiciones sobre el lugar que le corresponde a la evaluación de programas en el *cambio social*.

Las resoluciones provisionales de estos dilemas, esbozadas en los párrafos que siguen, ilustran que la evaluación cualitativa no se debe entender como una opción pragmática y de «caballo según carrera» que interviene detrás de una investigación cuantitativa de mayor

prestigio reconocido, tal como con demasiada frecuencia favorece la clase dirigente de la evaluación.

Procesos, resultados y racionalidad

David Campbell es el decano de los experimentos de campo aleatorios y de los cuasi-experimentos. En el capítulo 2 veíamos que sus primeros trabajos, elaborados a través de sus sucesivas exposiciones de las amenazas a la validez interna de tales estudios, supusieron el punto de referencia para los futuros trabajos de análisis de causas en los trabajos de campo. Decía el autor que de éstos se podían hacer inferencias causales más ciertas. Desacreditó la idea de una aplicabilidad limitada de las tareas aleatorias y aceptaba las consecuencias éticas de considerar la escasez de recursos como si de una lotería se tratara. Sostenía que el mayor obstáculo era la voluntad política.

Pero en la obra de Campbell se encuentran otros pasajes y mucho más ambivalentes. Reconoce que:

- Los sistemas de medición tienen una rigidez precaria.
- Muchas veces se ignora el proceso, por lo que, incluso cuando se descubren los efectos del programa, tal conocimiento tiene, en el mejor de los casos, unas implicaciones equívocas para la mejora del programa.
- Durante la evaluación, son casi inevitables los cambios en el programa de intervención.
- La experimentación es tediosa y equívoca, incluso en situaciones bien controladas.
- La utilidad política de los experimentos está limitada por la lentitud con que éstos se completan.

Conclusiones similares se sacaron de experimentos de campo realizados en Gran Bretaña durante la década de 1960 y principios de la de 1970 (Brody, 1976; Clarke y Cornish, 1972).

Campbell pretendía mejorar los diseños de la evaluación, en cambio Cronbach adoptó una postura más radical. Como decíamos en el capítulo 2, en la base de esta postura está la idea de que en la «validez externa», la validez de las inferencias que trascienden de los datos, está el *quid* de la acción social, no en la «validez interna» (Cronbach y otros, 1980, pág. 231). A partir de este análisis, Cronbach decide el orden de prioridad de la interpretación y la explicación de

los mecanismos que operan en un contexto local, con el fin de poder hacer inferencias verosímiles sobre otras situaciones, otras personas y otras intervenciones que tengan interés para los responsables de las políticas. Lo que importa es la extrapolación, «una predicción sobre qué se va a observar en un estudio cuyos sujetos u operaciones se alejen en algún sentido del estudio original» (Cronbach, 1986, pág. 94). Tiene cuidado de no rechazar nunca la defensa lógica de los experimentos comparativos con tareas controladas cuando el evaluador necesita abordar unas preguntas sumativas y causales, pero margina, sin embargo, este tipo de investigación porque piensa que la validez interna «no tiene una gran importancia en la evaluación» (Cronbach y otros, 1980, pág. 314).

En los últimos diez años se ha producido un cierto renacimiento de la fe en la evaluación de eficacia basada en los resultados. Una situación que ha estado favorecida por el compromiso político de algunos gobiernos occidentales con los planteamientos de la gestión eficaz, la rendición de cuentas del sector público y la adopción de indicadores de rendimiento para los centros educativos, los hospitales, los organismos de asistencia social y toda la variedad de servicios públicos. Baste decir que no estoy convencido de que la nueva generación de evaluadores que se basan en las pruebas haya resuelto adecuadamente los problemas que Campbell y Cronbach plantearon. En algunos casos, ni siquiera se han reconocido.

Las ideas racionalistas de eficacia no se ajustan a las ideas de participación democrática. Cronbach llega a decir que «el racionalismo se acerca de forma muy peligrosa al totalitarismo» (Cronbach y otros, 1980, pág. 95). En respuesta a las tesis de que los diseños cuasi-experimentales son convincentes y verosímiles, él y sus colegas dicen:

La exigencia de que una argumentación ha de ser «convinciente» no es más que un intento racionalista de definir una categoría de conclusiones que en modo alguno se basan en la creencia y en los valores; en la evaluación es un intento de situar las decisiones fuera de la política, estableciendo para ello una conclusión inevitable (1980, pág. 292).

Existe otro problema al que se presta poca atención. La pregunta «¿Funciona?» es una pregunta escéptica y «actúa como un guardián excluyente» (Bogdan y Taylor, 1994, pág. 296). Incluso suponiendo que los problemas de diseño se pudieran resolver, la pregunta seguiría careciendo de utilidad para los profesionales: «Los profesionales conscientes no se plantean su trabajo desde el escepticismo; creen en

lo que hacen» (1994, pág. 297). Bogdan y Taylor se han dedicado a desarrollar lo que ellos llaman una «investigación optimista». «Hemos desarrollado un enfoque de la investigación que nos ha ayudado a salvar la brecha que existe entre los activistas, por un lado, y los investigadores escépticos de base empírica, por otro» (1994, pág. 295). Sus estudios se han centrado principalmente en cómo se pueden integrar en la comunidad las personas con discapacidades graves. En vez de preguntar si los servicios son eficaces, preguntan: «¿Qué significa la integración?» y «¿cómo se puede conseguir la integración?».

Teorizar sobre los programas

Si tomamos la sociología como ejemplo, la teoría puede significar tanto un esquema sociológico (por ejemplo, la «teoría feminista») como unos conocimientos teóricos sustantivos de carácter sociológico (por ejemplo, el riesgo, o el estigma). Greene (1993, 1994) tal vez sea la defensora del lugar que la teoría sustantiva debe ocupar en la evaluación cualitativa de programas que más se hace oír. En los planteamientos interpretativistas de la evaluación, la *naturaleza* y la *forma* de la teoría serán específicas del contexto, con múltiples significados que surgen de las representaciones del programa. Los *principios* para la generación de la teoría deben evitar contentarse simplemente con comprender y deben fomentar la responsabilidad científica. Sobre la pregunta de cómo hay que incorporar esta teoría a la evaluación, Greene afirma:

El reto que se plantea es incorporar a la evaluación cualitativa los temas relacionados con ella, informar estos temas mediante perspectivas e interpretaciones locales y contextualizadas y, de este modo, aportar a la ciencia una teoría interpretativa de programas e incluir una voz interpretativa en el ámbito político más amplio (1993, pág. 35).

Y esto, en su opinión, se puede hacer de cuatro formas.

La primera implica explicar las disposiciones teóricas del evaluador —sus creencias, sus presentimientos, sus anteriores experiencias y conocimientos sobre esos temas—, que después se deberán vincular a un esquema de la teoría del programa. Es algo parecido a la recomendación inestimable de McCracken de realizar una revisión cultural (McCracken, 1988; Shaw, 1997). La segunda forma es mediante la descripción de las teorías de los programas locales: la «teoría tal como

se desea». De esta forma se dará una perspectiva teórica localmente significativa a la interpretación de los datos y se podrá comprender la forma de reaccionar ante las recomendaciones futuras. La tercera forma consiste en atender a los temas teóricos nuevos. En este sentido, la actitud abierta de las posturas cualitativas «a menudo es difusa y de objetivos poco claros, como si se tratara de unas antenas que van girando en busca de alguna señal» (Greene, 1993, pág. 38). Por último, la cuarta forma es integrar la teoría del programa en las conclusiones y las recomendaciones de la evaluación. De este modo, Greene desea que los evaluadores se muevan entre la postura etnográfica inductivista clásica y el teorizar apriorístico. En resumen, la autora propugna «para el evaluador un papel más proactivo y de sensibilidad ante la teoría» y «para la ciencia social interpretativa un papel más sólido, más reivindicativo y más persuasivo» (1993, pág. 40). «Al aceptar este reto, las historias de los evaluadores cualitativos pueden adquirir una mayor receptividad y mejorar su autoridad» (1993, pág. 41).

En una nota final de ese artículo, Greene reconoce que la reacción de los evaluadores críticos será una «receptividad ampliada», pero ¿hacia quién? y una «mejora de la autoridad», pero ¿con qué fin? Ha intentado desarrollar una respuesta a estas preguntas en trabajos posteriores, cuando dice:

Sin una postura comprometida y responsable como la de la ciudadanía científica, me temo que la evaluación cualitativa quedará marginada de los debates sobre los acuciantes temas públicos [...] Ya no podemos ocultar nuestra condición de ciudadanos tras nuestras subjetividades científicas. Debemos convertirnos en ciudadanos científicos (1996, págs. 286-287).

Los teóricos críticos quizá querrán ir más allá, como lo ilustra la obra de Carr y Kemmis en el ejemplo 5.2.

EJEMPLO 5.2. TEORÍAS CRÍTICAS DE LOS PROGRAMAS

Uno de los argumentos más extendidos a favor de la definición de la relación entre teoría y práctica en la evaluación es el de Carr y Kemmis (1986). Los problemas prácticos, dicen estos autores, siempre se solucionan con la acción y nunca con el saber, y la prueba de la investigación educativa es si resuelve los problemas educativos y mejora la práctica. Pero la teoría y la práctica son tareas «prácticas». La teoría como teoría sustantiva sobre algo siempre deriva de la teoría en el sentido de

un esquema. Del mismo modo, la práctica deriva de la teoría en el sentido de un esquema. Así pues, en este sentido los problemas educativos son siempre problemas de relación entre la teoría y la práctica.

Por consiguiente, cuando las relaciones entre teoría y práctica se entienden como el fracaso de los profesionales en aplicar las teorías que han desarrollado quienes se dedican a actividades teóricas, se distorsiona la realidad. Considerar que los problemas de la teoría y la práctica son una ruptura de la comunicación que afecta a los profesionales significa que no se sabe ver que los problemas prácticos de este tipo se producen en el transcurso de cualquier quehacer teórico. Presumir que las dificultades de esta índole de algún modo se pueden identificar y abordar en teoría, y luego «aplicarlas» a la práctica, suele ocultar el hecho de que se generan en la propia práctica. Además, convertir los conocimientos teóricos en normas de acción supone ignorar que son los profesionales quienes resuelven los problemas prácticos. La teoría que guía su práctica, y no cierta teoría ajena a ellos, es la que ofrece los principios rectores. Carr y Kemmis resumen así los defectos de los enfoques positivista e interpretativo:

El planteamiento científico, al ignorar que los problemas educativos siempre están preinterpretados, elimina efectivamente su carácter *educativo*. El planteamiento interpretativo, al aislar las interpretaciones que los profesionales hacen de su actuación de la crítica directa, concreta y práctica, elimina efectivamente su carácter *problemático* (1986, págs. 117-118).

En respuesta a esta crítica, los autores desarrollan un enfoque crítico de la teoría y la práctica, basado en la obra de Habermas y practicado como una ciencia participativa de aquellos que «crean, mantienen, disfrutan y soportan» las disposiciones educativas (1986, pág. 158).

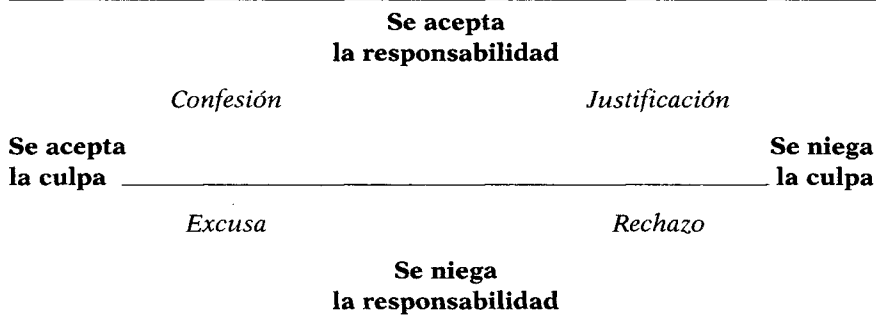
Las posturas críticas plantean unos problemas que tienen su origen en la teorización apriorística y en la tendencia a proclamar la posibilidad de una evaluación crítica sin, al mismo tiempo, ejemplificarla. Pero si la evaluación no combina la teoría y el propósito normativo común, seguirá careciendo de dirección o visión. Una pregunta que plantea el propósito normativo es si los compromisos con la justicia social deben formar parte de la evaluación. Volveremos a esta cuestión en el capítulo 7.

Evaluación y rendición de cuentas

Chelimsky (1997) sugiere tres esquemas conceptuales para los evaluadores: la evaluación para rendir cuentas, la evaluación para avanzar y la evaluación para conocer. Cronbach cree que hay que juzgar la evaluación «por su contribución al pensamiento público y a la calidad del servicio que se preste con posterioridad a la evaluación» (Cronbach y otros, 1980, pág. 64), por esto considera que la rendición de cuentas es una visión limitada de la evaluación de programas. «Creemos que el mejor uso de la evaluación no es someter a presión a los funcionarios públicos» (1980, pág. 17). «Con demasiada frecuencia, culpar a las personas se convierte en el principal uso de la rendición de cuentas, al tiempo que se olvida la mejora del sistema» (1980, pág. 135). Así, la evaluación recae con fuerza en la persona equivocada: «A quienes mayores responsabilidades se exige es a esos funcionarios públicos condenados a cultivar en pedregales, en unas condiciones climatológicas caprichosas» (1980, pág. 137). La conclusión del autor es que pedir responsabilidades es signo de patología del sistema social: «Tal exigencia, siempre que se ha producido en los últimos cien años, ha sido una señal de descontento: se cree que los responsables de los servicios son ineficientes, insuficientemente honestos y no autocríticos» (1980, pág. 139).

Sin embargo, dado que quienes ostentan la autoridad política o de gestión son quienes nombran a los evaluadores, «la imagen dominante de la evaluación sigue siendo “retributiva”» (Eraut, 1984, pág. 36). Los evaluadores cualitativos de programas desean tratar el tema de la responsabilidad de forma más problemática. En este sentido, el análisis de Lyman y Scott es útil. Distinguen dos tipos de rendición de cuentas: las «excusas» y la «justificación» (Lyman y Scott, 1970). Se da una excusa cuando alguien acepta que una acción es mala, errónea o inapropiada, pero no acepta la responsabilidad. En la justificación se acepta la responsabilidad, pero se niega la cualidad peyorativa de la acción. Esta distinción básica se puede matizar más de la forma que apreciamos en la figura 5.1 (pág. siguiente).

A los profesionales de la asistencia social, por ejemplo, parece preocuparles la responsabilidad o rendición de cuentas. En los estudios se han observado casos recurrentes de prácticas que los asistentes sociales consideran que no han ido bien y que pueden provocar cierta inseguridad en el profesional reflexivo (Shaw y Shaw, 1997a). Estos estudios indican que existía una sensación muy extendida de que los resultados, negativos y positivos, no se podían prever y nunca deja-



Fuente: Bull y Shaw (1992, pág. 641).

FIGURA 5.1. *Conceptualización de las responsabilidades.*

rían de aparecer y sorprender al profesional. En segundo lugar, los asistentes sociales dudan de que el trabajo «bueno» merezca un reconocimiento y que en el «malo» haya que buscar culpables y exigir responsabilidades. Las dos realidades, la predicción y la rendición de cuentas, están unidas. En palabras de un profesional:

Creo que, si tenemos en cuenta el hecho de que no se puede prever qué van a hacer los jueces en los grandes casos, ni qué es lo que harán en casos menores, en cierto sentido no se puede aceptar la culpa de un fracaso. Si no se puede aceptar la culpa si algo no funciona, ¿cómo se puede aceptar el mérito si funciona? (citado en Shaw y Shaw, 1997a, pág. 77).

Las conceptualizaciones y las pruebas empíricas de este tipo abren una fructífera perspectiva para los estudios de evaluación cualitativa de programas (Bull y Shaw, 1992).

Evaluación y cambio social

Ya apuntamos antes este tema. Los intercambios dentro de la evaluación cualitativa entre los evaluadores comprometidos con un enfoque interpretativista o naturalista, y aquellos que propugnan diversas versiones de la teoría crítica, significan que las consideraciones sobre el cambio social afloran permanentemente pero siguen siendo una fuente de desacuerdo. Por ejemplo, los evaluadores interpretativos suelen defender una agenda de acción social. Sin embargo,

en consonancia con el interpretativismo, los perfiles y las facetas concretos de esa acción no se prescriben, sino que surgen de la situación. De esta forma, el constructivismo difiere de los planes más prescriptivos de la potenciación, la igualdad y la justicia social de los planteamientos normativos críticos, feministas y otros (Greene, 1994, pág. 540).

En este sentido, Greene critica las posturas interpretativas. Sólo le convence en parte la confusión de las fronteras que separan las formas de conocer que ofrecen la ciencia social, la literatura y otras ramas de las humanidades:

El investigador social aplicado, a diferencia del escritor, el artista y el historiador, es responsable de cómo se lean y se interpreten sus historias, de las acciones que a partir de ellas se emprendan, y de quién se beneficie y quién no como resultado de la investigación. La investigación social responsable debe asumir sus consecuencias como parte de su propósito (Greene, 1992, pág. 42).

El interpretativismo justifica los valores, pero no justifica ningún valor determinado. En el fondo sigue siendo pluralista. La conclusión de Greene es que, a menos que se pueda modificar este relativismo esencial, «la indagación interpretativa siempre se juzgará con criterios que irremediabilmente reflejan las intenciones y los propósitos del juez» (1992, pág. 43).

EVALUACIÓN DE POLÍTICAS Y USO DE LA INFORMACIÓN

La evaluación es mucho más que una evaluación de programas. Es *muchísimo* más que la evaluación de los *resultados* de un programa. Ya hemos visto que la obra de Weiss, y de otros autores que abordaron problemas similares, cambió el rostro de la evaluación en los años setenta. La autora se ocupó de tres temas relacionados. Primero, *perfiló el contexto político en que se sitúa la evaluación*. Aunque su interés primordial está en la investigación de evaluación y de políticas en el ámbito federal, sus trabajos empíricos con las personas del campo de la política y de los programas resuenan en toda la teoría y la práctica de la evaluación. Segundo, expuso las limitaciones de las ideas instrumentales convencionales del uso político de la información, mediante su conceptualización del uso como ilustración. Por razones similares, Finch critica los modelos de uso de la evaluación

en los que priman el conocimiento y la resolución de problemas, porque se basan en un modelo racionalista:

El modelo racionalista del quehacer político entiende éste como una serie de sucesos discretos, donde se define claramente todo tema que haya que decidir, y las decisiones las toma un grupo concreto de actores que escogen entre unas opciones bien definidas, después de sopesar las pruebas sobre los resultados previsibles de cada una de ellas (1986, págs. 149-150).

El modelo de ilustración, o explicación, «ofrece mucho más espacio a la investigación cualitativa, a través de su insistencia en la comprensión y la conceptualización, más que en proporcionar unos hechos objetivos» (1986, pág. 154).

En tercer lugar, Weiss imbuó los modelos de uso de una idea realista del interés público: «Más que cualquier otra cosa, [Weiss] ha luchado por una teoría realista del uso. Estos cambios iniciaron en el mundo de la evaluación un debate, que sigue vivo en la actualidad, sobre el papel del idealismo y el pragmatismo en la evaluación» (Shadish y otros, 1990, págs. 207-208).

Las cuestiones referentes al uso de la información de la evaluación son fundamentales en cualquier análisis de la evaluación de políticas. Con palabras que recuerdan los poemas de Ronald Laing, Chelimsky capta las complejidades del uso de la información en una serie de aforismos:

- No sabes, pero no sabes que no sabes: por eso te arriesgas.
- No sabes, pero *sabes* que no sabes: por eso eres un poco más prudente.
- Sabes, pero no sabes realmente que sabes: por eso dudas.
- Sabes y sabes que sabes (1995, pág. 99).

Si la toma de decisiones en la política y en las comunidades políticas es impredecible y contingente, también se vislumbra de diversas formas en el ámbito de la evaluación. Por ejemplo, la evaluación que la etnografía emancipadora y crítica fomenta tiene una idea del cambio social, del uso de la información y de la relación de la política con la evaluación muy distinta de los supuestos implícitos en los modelos de evaluación orientados al descubrimiento, constructivistas y de ilustración, del tipo que se asocia con Stake y otros.

Sin embargo, las ideas sobre el uso de la información que desempeñan un papel fundamental para la ilustración cuentan con un am-

plio respaldo general. Cronbach y sus colegas apoyan este análisis con sus trabajos sobre el contexto del gobierno y sus ideas sobre una comunidad que es la que configura la política. En el capítulo 2 señalábamos que Cronbach y sus colegas sostienen que la evaluación entra en un contexto de gobierno típicamente de *acuerdo*, no de *mando*. Lamentan que la teoría de la evaluación se haya desarrollado «casi por completo en torno a la imagen del mando», con la idea asociada de unos gestores y unos funcionarios políticos que rigen con mano dura los controles de la toma de decisiones. Ellos piensan, por el contrario, que «la mayor parte de las acciones están determinadas por una comunidad plural, y no por la decisión de una sola persona» (Cronbach y otros, 1980, pág. 84).

En una versión modificada de las tesis de Cronbach, nosotros decimos que la evaluación está destinada a los funcionarios públicos y al público en general. Entre los primeros se encuentran los cargos electos, los funcionarios de las políticas, los funcionarios responsables de los programas y el personal ejecutivo, por ejemplo, los profesores, los asistentes sociales, el personal sanitario y los directores de residencias de ancianos. Entre el público relevante de un programa están la clientela del programa, y aquellos que Cronbach denomina iluminadores (por ejemplo, periodistas, académicos, algunos novelistas y dramaturgos, o comentaristas de los medios de comunicación). También podríamos incluir a los principales grupos de presión, cuya función, en algunos casos, se puede acercar a la de los funcionarios públicos (por ejemplo, los comités asesores o las organizaciones no gubernamentales) y, en otros casos, puede estar más asociada a los iluminadores (como en el ejemplo de los grupos de presión que se ocupan de los indigentes). Es importante reconocer que la comunidad que configura la política se dilata y se contrae de acuerdo con los temas y las fases de cualquier programa.

No obstante, no hay que exagerar el énfasis puesto en la importancia de las fases de desarrollo y madurez del programa. Podría significar una idea demasiado racional de los programas y los proyectos. Además, no refleja la realidad política actual. Muchos programas nunca pasan a un estadio de programa oficial y, desde el primer día, funcionan con un presupuesto fijo. Muchos de los estudios de programas sobre la pobreza de Gran Bretaña y Estados Unidos fueron de este tipo. El apoyo oficial a los programas y los proyectos innovadores muchas veces se condiciona a que el espíritu, los criterios y las prácticas innovadoras de un determinado «prototipo» impregnen los servicios existentes, y a que la responsabilidad financiera pase de la

administración central a los gobiernos locales o estatales cuando concluya la fase «piloto».

Ha habido cierta tendencia a tratar de forma simplista el análisis de Weiss y a suponer que todo uso de la información se hace a través de una ilustración gradual. Chelimsky realiza unas observaciones de mucho valor a partir de su larga experiencia en el que quizá sea el cargo más influyente en el mundo de la evaluación: el de Interventor General Adjunto de la Evaluación y la Metodología de Programas, de la General Accounting Office (GAO) de Estados Unidos. Dice, por ejemplo, que el uso de la evaluación puede operar mediante una función de *disuasión*: «En otras palabras, la mera presencia de la función, y la probabilidad de una evaluación persuasiva, puede evitar o detener una gran cantidad de prácticas indeseables del gobierno» (Chelimsky, 1997, pág. 105). Podemos darle la vuelta a la moneda y decir que el resultado indirecto de la existencia de una evaluación también puede ser un estímulo más positivo aunque más general a la práctica responsable y reflexiva.

Añade Chelimsky: «Y lo que es más importante [...] ocurre muchas veces que tanto la evaluación de rendir cuentas como la de conocer se emprenden *sin confiar en que se puedan utilizar*» (1997, pág. 105). Esta no utilidad esperada es característica de algunas de las mejores evaluaciones, incluidas «las que cuestionan las creencias populares y extendidas de unos tiempos de ideologías, o amenazan los intereses poderosos y afianzados en cualquier momento» (1997, pág. 105). De modo que «hay algunas muy buenas razones de por qué las evaluaciones pueden ser especializadas, y al mismo tiempo no se usen» (1997, pág. 105). Los comentarios de Chelimsky son a la vez sensatos y verosímiles:

Justificar todas las evaluaciones con un solo tipo de uso es una idea limitadora más que capacitadora, porque empuja a los evaluadores hacia una preocupación excesiva por que sus conclusiones sean aceptables para los usuarios, y plantea el peligro de convertir las evaluaciones en unas reiteraciones banales del *statu quo* (1997, pág. 106).

Este análisis no se debe emplear para hacer borrón y cuenta nueva, olvidar la obra de Weiss y empezar de nuevo con una idea sencilla y *ad hoc* del uso de la información. No nos debemos dejar persuadir, por ejemplo, por las ideas racionalistas en las que Patton (1988a) basa su defensa de una teoría instrumentalista del uso. Para reconfigurar las conclusiones que sacábamos en el capítulo anterior, es previsible que las explicaciones de ilustración e instrumentalistas del uso se asocien con una o más características afines de la evaluación (véase la tabla 5.2).

TABLA 5.2. *Modelos de evaluación y uso de la información.*

Modelo de uso	Objetivo de la evaluación	Base de la evaluación	Vínculos disciplinares	Espacio de tiempo para la acción
<i>De Ilustración</i>	Evaluación de políticas Desarrollo de programas Evaluación de defensa	Enseñanza superior	Vínculos disciplinares más fuertes Dedicación a la teorización	A más largo plazo
<i>Instrumental</i>	Evaluación de proyectos Estudios de viabilidad de programas Evaluación de profesionales	Evaluación «desde dentro» patrocinada por la propia organización Autoevaluación	Teorización limitada de las ciencias sociales	«Aplicaciones inmediatas»

No hay que interpretar de forma excesivamente estricta las conexiones que se muestran en la tabla 5.2. Ésta indica las tendencias probables con las que es más compatible una determinada idea del uso. No hay que leerla, por ejemplo, para concluir que la teorización está fuera de lugar en la evaluación del profesional, o que la evaluación de defensa siempre se produce en la enseñanza superior y tiene unos planes de acción a largo plazo.

El análisis que Chelimsky hace de algunas de las paradojas del uso de la información sugiere una idea correspondiente respecto a las influencias negativas y positivas del proceso político. Ya sabemos de qué forma los procesos políticos pueden actuar negativamente. Entre otras cosas, pueden favorecer:

- Las presiones para limitar el alcance de la evaluación.
- La exigencia de que los evaluadores se ajusten a unos esquemas temporales irreales.
- La presión indirecta para distorsionar los resultados del estudio, pidiendo para ello que se consideren las interpretaciones alternativas de los datos.
- La difusión selectiva de los resultados de la evaluación.

- La supresión o el retraso decisivo de la publicación del informe de la evaluación.

De todo ello se podrían contar muchas anécdotas en una imaginaria fiesta de evaluadores.

No obstante, Hedrick, que señala todas estas influencias negativas, además de otras, sugiere, a partir de su experiencia en la GAO, de qué forma la política puede ofrecer un respaldo paradójico a la evaluación. Su idea básica es que la evaluación a menudo supone una solución al desacuerdo político. Esto, aunque en sí mismo puede ser una consecuencia negativa, puede tener unos efectos positivos. Por ejemplo, los desacuerdos políticos pueden servir de estímulo para que se inicien estudios evaluativos:

Aunque sería tranquilizador pensar que los estudios de evaluación se realizan principalmente para averiguar si un programa funciona, o qué cambio político concreto ha propiciado, lo más probable es que muchos estudios se inicien para confirmar las creencias existentes o una postura política (Hedrick, 1988, pág. 9).

Además, el uso político de la evaluación como táctica dilatoria puede producir el resultado positivo de que «se puede ampliar la variedad de preguntas que se deban considerar, y los investigadores pueden adoptar una perspectiva a largo plazo e ir añadiendo conocimientos sobre un programa o un problema sociales» (1988, pág. 9). En otras palabras, las tareas de extinción políticas destinadas a limitar la utilización instrumental directa de la evaluación pueden alimentar sin darse cuenta los usos de ilustración y, a más largo plazo, de la evaluación.

Asimismo, «los desacuerdos políticos, por su propia naturaleza de confrontación, pueden ser un factor importante que propicie la existencia de entidades que realicen estudios evaluativos» (1988, pág. 9). En casos concretos, tales desacuerdos pueden servir para aumentar la visibilidad de los resultados de los estudios evaluativos. En este sentido, Chelimsky cita un caso interesante de una polémica que siguió a una síntesis de estudios de la GAO sobre la guerra química para el Departamento de Defensa (Chelimsky, 1995).

EVALUACIÓN DE POLÍTICAS Y POLÍTICA SOCIAL

¿Todo este análisis de los modelos de uso ha marcado alguna diferencia en la aportación que la investigación de políticas y, en particular, la investigación cualitativa de políticas, hace a la política pública y social? Hay un comentarista que piensa que no, y señala que parte de la responsabilidad de tal estado de las cosas está en la comunidad académica: «En las ciencias sociales actuales no existe una tradición sólida de amplia base de aplicar el estudio cualitativo específicamente a asuntos políticos» (Rist, 1994, pág. 556). Es una conclusión demasiado negativa y que no cuenta con el aval suficiente de los estudios y la experiencia de fuera de Estados Unidos. No obstante, los trabajos realizados en Gran Bretaña indican que hay pocas razones para sentirse satisfechos.

Los investigadores ingleses de políticas que trabajan con el modelo cualitativo han realizado unos análisis comparables de la relación entre conocimientos y política. Bulmer, por ejemplo, distingue entre los modelos empíricos, de ingeniería y de ilustración. El *empirismo* conlleva «una concepción de la investigación social que implica la producción de datos exactos —meticulosos, precisos, generalizables— y los propios datos constituyen un fin de la investigación» (Bulmer, 1982, pág. 31). El trabajo del investigador social consiste en producir unos hechos que se puedan usar. El autor asocia esta idea con la tradición que, pasando por Townsend y Donnison, se remonta a Titmuss y luego a Rowntree y Booth.

El modelo de *ingeniería* opera a la vez con una clara distinción entre investigación básica y aplicada, y con una relación explícita entre cliente y contratista para la investigación y la evaluación aplicadas. Se define el problema, se identifican los conocimientos de que se carece, se abre el concurso de propuestas de evaluación y se encarga el estudio. El problema se define fuera de cualquier disciplina concreta de las ciencias sociales, y la investigación resultante es un producto para la comunidad política, no para la disciplina académica. Tanto en Gran Bretaña como en Estados Unidos se han desarrollado muchas investigaciones evaluativas y teorías en torno a este modelo. En Gran Bretaña tiene sus orígenes en el muy influyente Informe Rothschild (Rothschild, 1971). Bulmer se queja de que el modelo de ingeniería subestima el hecho de que las fronteras entre investigación pura y aplicada están muy desdibujadas, y convierte al investigador en un técnico. Defiende un modelo de *ilustración* en términos casi idénticos a los de Weiss. Una posible consecuencia importante de un mo-

delo de ilustración es que parece justificar que se dé mayor prioridad a la investigación básica que en última instancia pueda demostrar ser útil —la investigación «estratégica», para emplear el lenguaje de los consejos de investigación del gobierno británico.

Finch (1986) da tres razones de por qué la investigación y la evaluación cualitativas tuvieron poco impacto en la política educativa y social de Gran Bretaña. En primer lugar, la utilidad de la estadística y de los «hechos» políticamente neutrales desde la perspectiva de los gobiernos da un atractivo engañoso a los informes estadísticos y basados en encuestas. En segundo lugar, el modelo fabiano de la relación entre el investigador y el responsable político ha sido el dominante. Sydney y Beatrice Webb formularon y consolidaron un modelo que quedó vinculado a la política del Partido Laborista. Influyó en los laboristas especialmente cuando estaban en la oposición y, en cierta medida, mientras estuvieron en el poder. Figuras como las de Halsey y Donnison formaron parte de una pequeña élite de académicos de elevado estatus, con unas posiciones políticas cómodas, que querían actuar como asesores y expertos profesionales para quienes estaban en el poder. Representaban un modelo de ingeniería social cuya finalidad era la justicia social a través de la fe fabiana en la «inevitabilidad de la gradualidad» y la investigación que era a la vez partidista y rigurosa. Esta tradición heredó el supuesto de la existencia de unos hechos políticamente no problemáticos que hablaban por sí mismos.

En tercer lugar, la administración social y la sociología se desarrollaron como disciplinas separadas en Gran Bretaña. Cuando la metodología cualitativa impregnaba la sociología, normalmente lo hacía sin un objetivo político. El resultado fue una ofensiva anticuantitativa en la sociología, y la tradición empírica de la administración social era uno de los objetivos de esta actitud. Hacia finales de los años setenta, en la sociología británica había una aversión prácticamente universal hacia la reforma social.

Desde entonces, esta posición ha cambiado, en parte debido a la aparición de las formas de investigación orientada a la acción del tipo que describimos en el capítulo siguiente. A mediados de los años ochenta, Finch se preguntaba si «el ambiente es hoy el adecuado para un grado de reinserción intelectual de los problemas sociales con los problemas sociológicos y para que la investigación cualitativa pueda hacer una aportación importante en este sentido» (Finch, 1985, pág. 112).

En la actualidad hay una aceptación gradual de las limitaciones de la investigación y la evaluación cuantitativas. Además del relativa-

mente modesto argumento de que existen unos aspectos complementarios en la investigación cualitativa y cuantitativa, la evaluación cualitativa de políticas es superior en tres sentidos. Primero, estudia los procesos y las acciones sociales en su contexto. Por ejemplo:

La dependencia de la investigación cuantitativa como base para comprender cómo se produce la conducta desviada en los centros educativos [...] será inevitablemente inadecuada, porque no puede considerarse de forma satisfactoria que los diferentes contextos producen diferentes resultados (Finch, 1986, págs. 165-166).

Segundo, estudia los procesos a lo largo del tiempo. Finch comenta sus tres años de estudios etnográficos con grupos de actividades lúdicas para niños de clase trabajadora y en edad de preescolar. Se preguntaba la autora: ¿Por qué fracasan estos grupos? Lo que interesa es que su trabajo aborda un tipo de pregunta que raramente se considera empíricamente en la evaluación (¿por qué fracasan los programas?). Además, era una pregunta que sólo se planteó durante la investigación cuando tres de los cinco grupos que la autora había escogido se deshicieron. Sin la dimensión longitudinal de la investigación cualitativa, esta pregunta evaluativa fundamental nunca se habría considerado.

Tercero, permite una investigación que cuente con una base teórica. Finch sostiene que una mezcla de teoría y datos es el distintivo de un buen trabajo cualitativo y que «esta particular mezcla produce precisamente el tipo de trabajo del que cabe esperar que va a tener un efecto en la política, porque ofrece unas reflexiones teóricas basadas en las pruebas» (1986, pág. 174).

Finch no elude las críticas que hablan de los defectos de la investigación cualitativa. Dice la autora que hay que reconocer que los responsables políticos educados según la tradición dominante siempre plantearán críticas, alegando que la investigación cualitativa es de una escala demasiado reducida, no representativa y, por consiguiente, no se puede generalizar y, además, con su insistencia en el significado, no consigue demostrar la correspondencia causal. Sugiere tres soluciones amplias. En primer lugar, hay que combinar los métodos cualitativos y cuantitativos. En segundo lugar, cuando se presenten las conclusiones, el investigador debe dar la información suficiente sobre la metodología para que el responsable de la política pueda valorarla, y demostrar que siempre se ha contado con el control de ésta. En tercer lugar, los investigadores cualitativos deben desarrollar

unos estudios comparativos para superar los defectos del estudio de casos singular y exclusivo. Recogemos de nuevo esta idea en el capítulo 7.

Los contextos políticos cambiantes hacen que las posturas de las que hemos hablado y que hemos recomendado en este capítulo tengan un carácter provisional. Cronbach y sus colegas señalaban con la ironía que da la experiencia que «habrá que revisar nuestras tesis para que se ajusten a los Estados Unidos del año 2000» (Cronbach y otros, 1980, pág. 14). Hammersley se muestra comprensivo con la idea de que las grandes expectativas respecto a una política que pueda influir son erróneas y que la investigación no es una fuente primera de consideración en el proceso político. A veces da la impresión de que así es como debe ser (Hammersley, 1994; 1995). El alineamiento provisional de Finch con un posición reformista no convence a Hammersley: «Cada vez es más difícil defender este optimismo de la ilustración» (Hammersley, 1994, pág. 146). Sin embargo, en el campo de la evaluación del profesional se producen avances que se basan en una lectura más optimista del valor potencial de la evaluación. Vamos a ocuparnos de ellos a continuación.

CAPÍTULO 6

LA EVALUACIÓN DEL PROFESIONAL

La evaluación es algo más que la evaluación de políticas y de programas. Su propósito, su alcance y su metodología se convierten en algo parcial y miope siempre que los profesionales de la evaluación sucumben a la tentación de malinterpretar su potencial. Stake fue quizás el primer teórico de la evolución que lo reconoció con su progresivo énfasis en la calidad de las actividades de evaluación, más que en los resultados; en la importancia de las interpretaciones tácitas e implícitas; en los públicos locales y básicos de la evaluación; en la «idea social antropológica» (1991, pág. 76); en la evaluación receptiva; y en los límites de los modelos de rendición de cuentas de la evaluación. Dio el estímulo necesario para la evaluación de proyectos y de profesionales, además de la evaluación de políticas y de programas.

Parte de la responsabilidad del fracaso de la evaluación de corte tradicional la sitúa en aquellos que ponen en primer plano la producción, la difusión y la utilización de los conocimientos: «Los investigadores han tenido escaso interés en estudiar los problemas entrelazados, personalistas y que se asemejan a una crisis de la práctica diaria» (Stake y Trumbull, 1982, pág. 4). Prefieren «amañar los modelos» y centrarse en cambiar los esfuerzos del aula y de la institución; en consecuencia, su investigación «no ha ofrecido a los profesionales las experiencias indirectas que podrían alterar las convicciones y las prácticas en los centros educativos» (1982, pág. 11).

Es en el ámbito de la evaluación cualitativa del profesional donde se han producido los avances más innovadores en los últimos años, y así seguirá siendo. El título de este capítulo es toda una premisa, por supuesto. Por «profesional» no nos referimos al evaluador, sino a los profesionales que prestan algún tipo de servicio en organizaciones, hospitales, centros educativos, organismos de asistencia social, etc. Pero ¿significa una evaluación *de* la práctica y de los profesionales realizada por investigadores cualitativos? ¿Una evaluación realizada *por* los profesionales? ¿Una evaluación que es *para* los pro-

fesionales más que para los investigadores académicos, los directores de programas o los responsables políticos? ¿Y qué decir acerca de la evaluación que se realiza sobre una base participativa *para* y *con* los padres, los niños, los pacientes y otros ciudadanos usuarios de los servicios públicos? En este capítulo se analizan estas distinciones. Consideraremos principalmente cuatro áreas de práctica de la evaluación, que se pueden subsumir en el amplio título de evaluación del profesional.

1. La investigación y la evaluación realizadas por los profesionales.
2. La investigación participativa.
3. La evaluación como una dimensión de la práctica directa, por ejemplo, mediante la indagación reflexiva.
4. La evaluación *para* los profesionales realizada normalmente por investigadores académicos.

PREGUNTAS PARA LA INDAGACIÓN DEL PROFESIONAL

Si recordamos lo que decíamos sobre los paradigmas en el capítulo 3, quedará claro que el hecho de adoptar una metodología cualitativa de evaluación no va a solucionar de la noche a la mañana los problemas que plantean las metodologías cuantitativas más tradicionales. En efecto, sin las debidas garantías, puede ser incluso que algunos de los problemas se agudicen más. Por ejemplo, McNamara se queja enérgicamente de que quienes entran en las aulas para realizar observaciones muchas veces «no saben comprender ni apreciar las realidades de la vida del aula tal como se ven desde la perspectiva del profesional. Adoptan la visión limitada y personal del intruso» (1980, pág. 114). «Utilizan sus experiencias de trabajo de campo para fomentar los objetivos de su propia evaluación» y «no sirven a los profesores que gustosamente les han permitido acceder a sus clases» (1980, pág. 115). Aunque McNamara no quiere entrar en las sutilezas de tal argumentación y sus tiros van dirigidos a mantener la investigación cualitativa y dejar al margen la evaluación del profesional, no está mal que los evaluadores oigan esas arremetidas tan sonoras contra sus dogmas.

En los capítulos 7-9 volveremos a los temas éticos que plantean estas críticas. Pero conviene señalar que ponen sobre la mesa dos problemas generales. Primero: ¿cómo debemos interpretar la relación entre la comprensión o indagación del profesional y la investi-

gación evaluativa? Segundo: ¿la comprensión tácita e implícita a la que Stake se refiere se opone a la indagación más explícita, planificada y cualitativa que realizan los profesionales?

La indagación y la evaluación del profesional

El planteamiento general del primer problema lo analizábamos en el capítulo 1, cuando preguntábamos si investigación y evaluación son más o menos la misma empresa o empresas diferentes. En la literatura existen abundantes argumentos opuestos según los cuales la evaluación y la práctica del profesional son o muy *diferentes* de la investigación evaluativa en general o muy *similares* a ella. Cochran-Smith y Lytle piensan que los principales estilos de investigación educativa, tanto los que se orientan a los resultados como los que se orientan a la comprensión de la «ecología del aula», no han tenido en cuenta la voz de los profesores, «las preguntas que los profesores plantean [...] los marcos interpretativos que emplean para comprender y mejorar sus propias prácticas en el aula» (Cochran-Smith y Lytle, 1990, pág. 2). Sostienen que la investigación del profesor (la «indagación sistemática e intencional realizada por los profesores»; 1990, pág. 3) no es una imitación de la investigación universitaria, sino que tiene «su propio género» (1990, pág. 4).

La moderada defensa que estos autores hacen de tal postura no siempre se refleja en otros defensores de la evaluación del profesional como un modo diferenciado. Las distinciones entre investigación evaluativa y evaluación del profesional se citan con demasiada frecuencia de forma interesada, para defender la «relevancia» del trabajo del profesional. Por ejemplo, Fortune (1994) sostiene que las diferencias entre la etnografía y la práctica de la asistencia social incluyen las siguientes:

- «La práctica se centra en la acción, mientras que los métodos cualitativos, incluida la etnografía, sólo pretenden describir» (1994, pág. 64).
- El profesional necesita una habilidad adicional para la lógica deductiva e inductiva, «además de un excelente sentido del tiempo para saber cuándo detener la recogida de datos» (1994, pág. 65).
- Los etnógrafos pretenden generalizar, mientras que en la práctica de la asistencia social «no existe una necesidad inherente de

comunicar esa realidad a otras personas ni de generalizar más allá de la experiencia de ese individuo» (ibíd.).

- Los estudios microetnográficos mejoran la práctica de la asistencia social y pueden «generar una información de una utilidad inmediata» (ibíd.), mientras que las macroetnografías requieren tiempo, son difíciles y «sólo indirectamente relevantes para la asistencia social» (1994, pág. 68).
- Dado que el tiempo y los intereses de los asistentes sociales se emplean en la práctica y en la acción, los estudios microetnográficos se ajustan bien a los recursos y las energías de que se dispone.

Esta forma de distinguir entre la práctica y la evaluación es casi completamente inútil. La defensa de unas destrezas distintivas del profesional, los supuestos referentes a la etnografía, la forma tan acrítica en que se presenta la generalización y la utilidad de la información, y las ingenuas aspiraciones del autor respecto a la posibilidad de que los profesionales dirijan unas microetnografías, llegan a un grado de simplicidad casi increíble. Las ideas de Fortune poco sirven para rebatir la tesis de Smith y Klein de que las distinciones que se basan en la epistemología o la metodología son «poco fundadas, cuando no insostenibles» (Smith y Klein, 1986, pág. 55). Hay que reconocer, en honor a la justicia, que tal defecto no se limita al ámbito de los profesionales. Cuando afecta a los investigadores académicos, la distinción contundente entre la indagación del profesional y la investigación evaluativa a veces se entiende como una defensa elitista de la complejidad y la superioridad conceptual de la metodología etnográfica (véase Atkinson y Delamont, 1993).

Sin embargo, existen diferencias importantes de *propósito* entre la investigación evaluativa y la indagación del profesional. Las cuestiones de los acuerdos de evaluación previos, los públicos previsibles y la utilización de los resultados plantean unos temas especiales para la evaluación del profesional. Por ejemplo, no es obvio que la importante obra de Weiss y de otros autores sobre los modelos de ilustración de la evaluación de políticas y programas se puedan trasladar perfectamente a la indagación del profesional, donde es previsible que el uso *inmediato* ocupe un lugar más destacado que el uso a largo plazo. También hay unas diferenciaciones importantes *dentro* de la investigación evaluativa, que de manera general se observan en las distinciones entre *investigación* evaluativa, *indagación* evaluativa y *razonamiento* evaluativo (véase Eisner, 1991a). Estas diferencias se

pueden analizar y comprobar si se piensa en los procesos de inferencia en que participan los diferentes profesionales e investigadores. El psicoterapeuta Erikson hizo aportaciones valiosas en este sentido, en un simposio sobre pruebas e inferencia en las ciencias humanas (Lerner, 1959). El autor lamenta que «los trabajadores clínicos muchas veces no saben hacer explícito, ni siquiera para sí mismos, qué inventarios de indicios de pruebas escrutan de forma regular aunque inconsciente» (Erikson, 1959, pág. 82).

Conocimientos tácitos e indagación cualitativa

La *segunda* pregunta es si la interpretación tácita e implícita se opone a un razonamiento más explícito, planificado y cualitativo del profesional. Parece que Stake se plantea lo mismo cuando con Trumbull dice que «para los profesionales [...] el conocimiento formal no es necesariamente un peldaño hacia una mejor práctica [...] Creemos que son los conocimientos personales los que guían en mayor medida la práctica» (Stake y Trumbull, 1982, pág. 5). Aunque no rechazan los conocimientos formales, «el punto de apoyo del cambio que muchas veces se olvida es la recogida disciplinada de conocimientos experienciales» (1982, págs. 8-9). Según esta idea, una buena evaluación del profesional desvelará unos conocimientos tácitos, «la comprensión en gran medida desarticulada y contextual que muchas veces se manifiesta en los asentimientos con la cabeza, en los silencios y en los comentarios humorísticos o satíricos» (Altheide y Johnson, 1994, pág. 492). Hay acciones, juicios y reconocimientos que se producen de forma espontánea. No tenemos que pensar en ellos antes de llevarlos a la práctica. Muchas veces no sabemos que hemos aprendido a realizarlos. Aunque alguna vez podamos ser conscientes de lo que necesitamos saber para realizar una acción, lo normal es que no sepamos describir los conocimientos que nuestra acción revela. Se ha convertido en un conocimiento de «pensar como de costumbre» (la expresión es de Schutz): «El conocimiento tácito existe desde el momento en que se realiza una acción que no se comprende, cuando la comprensión se ofrece sin articularla y cuando se llega a unas conclusiones sin argumentarlas» (Altheide y Johnson, 1994, pág. 492).

El problema que para la evaluación supone este conocimiento, en gran medida no discursivo, es «cómo hablar de aquello de lo que raramente se habla» (1994, pág. 493). Los evaluadores comparten con los investigadores la ambivalencia referente a esos conocimientos,

con una tendencia a oscilar entre considerarlos como potencialmente liberadores o restrictivos (véase Scott, 1990). Así se observa en algunos planteamientos feministas de la investigación evaluativa. Hawkesworth, por ejemplo, critica el llamamiento a la intuición que se hace en ciertas investigaciones de corte feminista, alegando que:

La intuición aporta una base para las afirmaciones acerca del mundo que es a la vez autoritaria, pues no admite más discusión, y relativista, ya que nadie puede refutar la aprehensión «inmediata» de la realidad de otra persona (Hawkesworth, 1989, pág. 545).

La mayoría de los teóricos de la evaluación probablemente comparten el rechazo de Hawkesworth de este tipo de llamamiento a la intuición y prefieren la interacción de algún tipo de «realidad racional» con la «realidad experiencial» (Stake, 1991, pág. 83). Esto se describe a veces como el ejercicio de la «sabiduría práctica» por parte de los profesionales. Por ejemplo, el influyente artículo de Schwab sobre la comprensión del currículo defiende el lenguaje de lo práctico, más que el de lo teórico. Por práctico «no entiendo [...] los objetivos familiares y de fácil consecución a los que se puede llegar con recursos habituales» (Schwab, 1969, pág. 1), sino «una disciplina compleja [...] que se ocupa de la decisión y la acción» (1969, págs. 1-2). Schwandt, partiendo del pensamiento aristotélico, prefiere una solución más radical. Distingue entre el conocimiento teórico («saber que»), el conocimiento artesano o de destrezas («saber cómo») y el conocimiento práctico-moral («saber a partir de»). El razonamiento ético tiene lugar a través de la sabiduría práctica. Schwandt pretende reformular la evaluación en conjunto dentro de este tipo de hermenéutica práctica: «La finalidad del conocimiento práctico-moral es mover de verdad a las personas, y no simplemente darles unas buenas ideas» (1997, pág. 81).

Este preámbulo tiene un mensaje obvio. Miremos dónde pisamos, evitemos soluciones precipitadas y mantengamos los propósitos evaluativos en primer plano.

LA INVESTIGACIÓN DEL PROFESIONAL

La idea del profesional como investigador no es nueva. El modelo de investigación clínica tiene unos precedentes honorables en la obra,

por ejemplo, de los médicos que ensayaban consigo mismos sus métodos de vacunación, y en la investigación clínica de la psicoterapia. No es nada nuevo en las profesiones más jóvenes. A finales del siglo XIX, Charles Loch dedicó mucha energía a defender que la asistencia social es como la ciencia: «Es la ciencia, la ciencia de la vida, en funcionamiento; el conocimiento realizando su trabajo perfecto» (citado en Timms, 1968, pág. 59). Lo que sí es nuevo es la diversidad de argumentaciones y de ejemplos que se ofrecen para la investigación del profesional, su aparición independiente en el mundo de la educación en Gran Bretaña y Estados Unidos en los años setenta, en la asistencia social norteamericana en esos mismos años y, de una forma diferente, en la asistencia social inglesa en los años ochenta, y el desarrollo gradual de las metodologías cualitativas. Consideraremos las diversas formas y argumentos de la idea del profesor como investigador, de la investigación clínica cualitativa, y de la investigación del profesional de la asistencia social.

El profesor como investigador

Los orígenes del movimiento del profesor como investigador en Estados Unidos se pueden buscar perfectamente en el gran interés por la investigación en la acción que, a través de autores como Lewin, se despertó en los años de la posguerra. Este movimiento estuvo al principio asociado a la investigación cuantitativa y al establecimiento de objetivos. La aparición paralela, aunque probablemente independiente, de los profesores como investigadores en Gran Bretaña giró sobre todo en torno a las grandes reformas curriculares de los años sesenta y setenta. En Gran Bretaña se ponía mayor énfasis en el proceso de la enseñanza que en el establecimiento de objetivos y en la investigación de la eficacia. Hammersley señala que el cambio hacia ideas progresistas sobre las relaciones entre alumno y profesor también alentó en los estudios curriculares el paso «de unas concepciones burocráticas de la enseñanza a otras profesionales, y de la idea del investigador como la persona que desarrolla [el currículo] a la de la persona que evalúa o facilita» (1993a, pág. 427). En otras palabras, los conceptos de «descubrimiento» e «indagación» surgieron en primer lugar en la educación y, de ahí, pasaron a la evaluación y a la investigación. Decíamos en el capítulo 2 que la obra de Stake se fusionó con los nuevos estudios del Centre for Applied Research in Education de la Universidad de East Anglia, en el Reino Unido.

Desde el principio hubo tendencias diferenciadas en esos trabajos. Una de las primeras tesis de gran influencia fue la de la evaluación iluminativa, debida a Parlett y Hamilton (1972). Se criticaba la evaluación clásica porque era artificial, porque «raramente a partir de una realidad “desordenada” se pueden generalizar unos resultados “ordenados”» (1972, pág. 5), y se defendía la evaluación de las innovaciones educativas en que «la pretensión de medir [...] los productos se abandona a favor de un estudio intensivo del programa en su conjunto» (1972, pág. i). Los autores distinguían entre el «sistema de instrucción» formal y el «medio de aprendizaje» informal, en el que se reinterpretan los sistemas de instrucción cada vez para un enclave determinado. De ahí que el enclave —el medio de aprendizaje— se convierta en el principal objetivo del interés de la investigación. En todo esto era esencial la idea de currículo oculto, por el que los alumnos asimilan las «convenciones, las creencias y los modelos de realidad que de forma constante e inevitable se transmiten a través del proceso de enseñanza» (1972, pág. 15).

En lo que a la planificación y la metodología respecta, «la evaluación iluminativa no es un paquete metodológico estándar, sino una estrategia general de investigación» (1972, pág. 15), basada en un proceso de atención progresiva mediante múltiples métodos, pero dando prioridad a una serie de métodos de observación. Parlett y Hamilton compendian su tesis en una buena analogía.

Para saber si una obra de teatro funciona, no hay que fijarse solamente en el texto, sino también en la representación; es decir, en la interpretación que de la obra hacen el director y los actores (1972, págs. 30-31).

La finalidad de la evaluación iluminativa es «estudiar la innovación a través del medio en que se representa» (1972, pág. 31).

Este traslado del peso hacia los métodos cualitativos ha sido más marcado en algunos representantes de este enfoque que en otros. En un artículo que se suele citar, escrito en los años setenta, Stenhouse, que fue importante en los primeros avances que se produjeron en Gran Bretaña, prefería el enfoque «antropológico social» que «por encima de todo [...] busca no sólo la generalización, sino también la caracterización de la exclusividad de las situaciones particulares» (1993, pág. 225). Algunos autores han ido más allá, y defienden que los profesores apliquen la etnometodología a la investigación del aula (Payne y Cuff, 1982).

Otros propugnan un eclecticismo metodológico. Kelly, por ejemplo, sostiene que la metodología concreta que se emplee es menos importante que el planteamiento general de la acción y la investigación, y está dispuesta a utilizar «datos duros» para apoyar sus ideas ante quienes no crean en los datos etnográficos: «Necesitan que la estadística y los porcentajes les convenzan de que no formamos parte de un grupo marginal de lunáticos, que sabemos de qué estamos hablando» (Kelly, 1985, pág. 145). Algunos llevan el eclecticismo al extremo de mostrarse escépticos ante cualquier reflexión metodológica. Bassey abre el capítulo que titula «Does Action Research Require Sophisticated Research Methods?» diciendo: «Me siento tentado a escribir el capítulo más breve de este libro, respondiendo a mi pregunta con una palabra: NO» (1986, pág. 18). En esta tradición, los énfasis se han puesto también en diferentes cuestiones en cada momento: en las relaciones con los grupos de interés, en el lugar que ocupa el proceso teorizador, en la epistemología y en el papel político de la evaluación (Simon, 1987, cap. 3).

La investigación clínica

La investigación clínica cualitativa no ha prosperado en el ámbito de la medicina. Se podría especular que tal vez se deba a las sólidas tradiciones cuantitativas de la investigación epidemiológica y a la fuerza de la investigación etnográfica de índole tradicional dentro de la sociología de la medicina. Sin embargo, cada vez es más la literatura que se ocupa de los métodos cualitativos en la atención a las personas mayores y en las profesiones paramédicas. En algunos casos se comparte el eclecticismo de ciertos evaluadores educativos, aunque Miller y Crabtree defienden con fuerza la investigación clínica cualitativa. Describen la mayor parte de la investigación clínica como «muchas conversaciones tras las paredes, pero un sufrimiento cada vez mayor en el mundo clínico», e instan a los investigadores clínicos cualitativos a «salir de sus escondites, participar de la experiencia crítica y de las cuestiones que plantea, y practicar la humildad y la fidelidad en el seno de una comunidad de discurso que está junto a esas paredes» (Miller y Crabtree, 1994, págs. 340, 349).

La mayor parte de las investigaciones clínicas que se han publicado consisten en ensayos y epidemiología clínicos que no consiguen dirigirse al «mundo real de la práctica clínica, que implica intenciones, significados, intersubjetividad, valores, conocimientos persona-

les y ética» (1994, pág. 341). Miller y Crabtree esbozan una estrategia triple:

- Crear un espacio abierto de investigación que fomente los enfoques cualitativos y críticos.

Proponemos que los investigadores clínicos estudien las cuestiones que plantea la experiencia clínica, atiendan a cualquier valor y supuesto subyacente y los desvelen, y dirijan el resultado hacia los participantes clínicos (1994, pág. 342).

- Ofrecer las herramientas necesarias para descubrir y confirmar las historias y los conocimientos clínicos.
- Identificar y describir los medios para compartir las historias y los conocimientos.

«El contexto local y la historia humana [...] son las metas fundamentales de la investigación cualitativa, y no la “generalizabilidad”», y las reglas de las pruebas «se pueden traducir para el público clínico a formas de contar unas historias que sean metodológica, clínica y retóricamente convincentes» (1994, pág. 348).

La investigación cualitativa del profesional de la asistencia social

La investigación del profesional de la asistencia social es diferente en Gran Bretaña y en Estados Unidos. En este último país existe un fuerte movimiento del «profesional científico» que se basa en general en la metodología cuantitativa y orientada a los resultados. La forma de evaluación de la que más se habla en el movimiento del profesional científico son los diseños de sistemas singulares. En las revistas de asistencia social hay un continuo debate sobre las aportaciones del movimiento de la práctica empírica. Los argumentos a favor de la investigación cualitativa del profesional provienen en parte de los observadores de la tradición de la práctica empírica. En el capítulo 2 señalábamos la aportación de Reid a esos avances (Reid, 1990; Sherman y Reid, 1994). Epstein dice que tanto la evaluación de corte tradicional como la práctica de la asistencia social han adoptado unas posiciones de «determinismo psíquico, que sitúan el origen del problema humano en el interior de la persona individual, en algún lugar no empírico, como el alma, el espíritu, la psique, la mente,

las estructuras de la personalidad y ámbitos similares» (Epstein, 1996, pág. 115). Aunque uno apela a la psicología científica y la otra a lo que ella llama psicología expresiva, ambos «se concentran en leer y cambiar las mentes de las personas pobres y de aquellos que padecen cualquier otro tipo de angustia o desviación» (1996, pág. 115). «El problema en ambos casos es que falta el contexto» (1996, pág. 116). La respuesta implícita al llamamiento de la autora parece incluir una estrategia de investigación del profesional más sociológica y cualitativa. Un enfoque en la construcción de tal estrategia ha consistido en analizar las diferencias en las estrategias de procesamiento de datos de la investigación de asistencia social y cualitativa, para valorar de qué forma se pueden poner en práctica.

Lang, por ejemplo, analiza las estrategias de recogida y procesamiento de datos de la metodología de la asistencia social y cualitativa, y recomienda que se adopte la última, no sólo para la construcción de conocimientos, sino también para «propósitos que deriven en acciones» (Lang, 1994, pág. 277). La autora cita como ejemplo el procesamiento de datos de los investigadores y profesionales cualitativos, e indica que «el profesional “ nombra ” los datos a través de la referencia a la teoría; el investigador “ nombra ” los datos a través de un proceso de conceptualización que deriva de las características de los datos» (1994, pág. 271). Cree que se plantean algunos problemas para el trabajo social porque «la presión por saber qué hacer, qué acciones emprender, puede cerrar el camino del desarrollo de los conocimientos a partir de la práctica para muchos profesionales» (1994, pág. 271). Propone un planteamiento de la práctica inductivo y de construcción de teorías: «La teoría existente debe tener un estatus más provisional, un lugar menos central en nuestra práctica docente, para abrir la posibilidad de construir teorías» (1994, pág. 276). Los asistentes sociales deben derivar sus acciones de las características de los datos, más que recurrir a la teoría existente como primer recurso. Ruckdeschel ha empezado a desarrollar unos métodos comparables para el estudio de casos. Es uno de los varios evaluadores en quienes ha influido la obra de Geertz (1973). Sigue la idea de Denzin de los «especímenes conductuales» como descripciones casi completas de las interacciones entre los individuos de un determinado marco temporal (Denzin, 1989), y acepta también que la tarea del estudio de casos es «dar voz al pobre» (Ruckdeschel y otros, 1994, pág. 255). Cree que en la asistencia social de Estados Unidos el escepticismo oficial ante los métodos cualitativos, que Kelly señalaba refiriéndose al mundo educativo, hoy queda ya muy en el pasado (Ruckdeschel, 1999). Hay ciertas difi-

cultades para aplicar esos modelos, pero son realmente prometedores para el desarrollo de una evaluación cualitativa del profesional.

La asistencia social ha tenido un desarrollo más pragmático en Gran Bretaña. Los principales teóricos de la investigación del profesional propugnan una metodología ecléctica, sin dar primacía a ningún enfoque concreto (Cheetham, 1998; Fuller y Petch, 1995). Los planteamientos de la práctica basada en pruebas están ejerciendo una influencia cada vez mayor a través de la obra de autores como Sheldon, Macdonald, Kazi y McGuire (sobre estos estudios véase Shaw, 1996a, cap. 9, y Shaw y Shaw, 1997b).

Los principales avances de la investigación cualitativa de la práctica en Gran Bretaña se han dado en el desarrollo de los enfoques de la investigación participativa y la indagación reflexiva, de los que nos ocupamos en las próximas páginas. Un desarrollo más amplio de unas ideas similares a las que proponía Lang se presenta en Shaw (Shaw, 1996a, 1997, 1998), quien defiende e ilustra los beneficios de «colonizar» y «traducir» la metodología cualitativa para la evaluación como una dimensión directa de la práctica.

La investigación crítica del profesional

El pragmatismo metodológico y la complacencia política de gran parte de la indagación del profesional ha fortalecido la postura de quienes piensan que se requiere una concepción más radical y crítica de la investigación del profesional. En el ámbito de la asistencia social, esto se ha manifestado sobre todo en la investigación participativa. Esta investigación tiene menos tradición en el ámbito de la enseñanza, donde dominan una tradición sólida de investigación que incluye el estudio de casos y la etnografía crítica. La obra de Carr y Kemmis, *Becoming Critical [Teoría crítica de la enseñanza]* (1986), probablemente es el ejemplo más influyente de estudio de casos crítico y de investigación en la acción.

La premisa de estos autores es que el movimiento del profesor como investigador surgió como producto de diversas presiones sociales, políticas y profesionales y, en consecuencia, ha sido oportunista y pragmático. Lamentan y rechazan las distinciones drásticas entre teoría y práctica, alegando que propician el dominio del lenguaje técnico: «La dimensión “moral” de la educación se elimina de forma inadvertida, y la educación se convierte en un asunto puramente técnico» (Carr y Kemmis, 1986, pág. 38). La estrategia docente debe proceder guiada

por la conciencia de que la educación está situada históricamente, es una actividad social e intrínsecamente política y, como tal, intrínsecamente problemática. Esos profesores se entregarán a la «autorreflexión crítica» (1986, pág. 40) y construirán oportunidades de discutir y debatir con otras personas ese discurso privado, para ayudar a establecer unas comunidades críticas de indagadores. De este modo se crea espacio para la investigación del profesor. En algunos momentos, esa investigación «sólo será una inquieta actitud indagadora sobre la enseñanza y el currículo» (1986, pág. 40), pero en otros casos en que el profesor adopta una «perspectiva de proyecto»:

las *acciones* que se emprendan se considerarán «provisionales» o «experimentales», se examinará de forma crítica el *lenguaje* con el que se describen y se interpretan las acciones, se analizarán y serán objeto de reflexión las *consecuencias sociales*, y se examinará la *situación* en que tiene lugar la acción para ver cómo crea y limita el potencial de la estrategia escogida (1986, pág. 40).

En resumen, dicen, «los profesores se convierten en radicales». Esto va a despertar de nuevo la disposición del juicio moral, y favorecerá una «ciencia educativa crítica» que «no es investigación *sobre* y *acerca de* la educación, es investigación *en* y *para* la educación» (1986, pág. 156). Tal ciencia debe propiciar la participación de todos aquellos que «crean, mantienen, disfrutan y soportan» las disposiciones educativas (1986, pág. 158) y transforman no sólo a los profesionales, sino también la propia educación.

Críticas etnográficas

El desarrollo de la investigación del profesional ha sido más o menos bien acogido en la asistencia social y en la atención a los mayores. Tanto en Estados Unidos como en Gran Bretaña se han financiado algunos proyectos de investigación del profesional en estos campos y en el ámbito de la enseñanza. Sin embargo, la acogida ha sido menos calurosa entre los etnógrafos de carácter tradicional. Algunos se contentan con estudiar los problemas metodológicos que el movimiento del profesor como investigador podría subestimar (Burgess, 1980; Burgess y otros, 1994). Otros, especialmente en Gran Bretaña, se han mostrado mucho más cáusticos en sus críticas, que implícitamente dicen que todo el movimiento no es más que una especie

de circo, basado en una evasión metodológica y una ingenuidad teórica (Atkinson y Delamont, 1993; Hammersley, 1992, 1993a). Por ejemplo, Atkinson y Delamont acusan a los investigadores de estudio de casos de «contentarse con invocar la antropología de forma ritual. Tiene una importancia totémica o fetichista» (1993, pág. 207). «La complejidad metodológica no es una marca característica del género» y sus profesionales son «antiintelectuales» y culpables de una «carencia de erudición» (1993, pág. 210). Acusan de primitivismo a la investigación que utiliza el estudio de casos —«un énfasis en la investigación como quehacer esencialmente vernacular o demótico» (1993, pág. 211)— y hablan de «la paradoja de un grupo de “expertos” en evaluación, parte de cuyos valores consiste en negar la experiencia» (1993, pág. 213).

Hammersley evita el desdén grosero y elitista y la simplificación exagerada que caracterizan a la obra de Atkinson y Delamont, y plantea preguntas más cuidadosas referentes a los supuestos ocultos del enfoque. En primer lugar, no cree que la enseñanza se deba considerar como «isomorfa de la indagación» (Hammersley, 1993a, pág. 441) y piensa también que existe el peligro de sobrestimar los beneficios que de la investigación se puedan conseguir: «Los productos de la indagación sistemática no serán necesariamente mejores que los supuestos integrados en las formas tradicionales de hacer las cosas. Es una falacia moderna presumir lo contrario» (1993a, pág. 438). Le preocupa que esto pueda significar una forma de imperialismo del profesor. Pese a su lenguaje igualitario, puede servir de estrategia profesionalizadora. En segundo lugar, cree que la crítica de la investigación convencional que se plantea desde el campo de la investigación democrática se basa en un concepto demasiado estrecho de la relevancia de la investigación y en una fe manifiestamente optimista en la capacidad de la investigación de influir en la política y en la práctica. Dice que «hay buenas razones para pensar que la investigación no puede resolver de forma rutinaria los problemas de los profesores» (1993a, pág. 430). «No existe un método científico que garantice los resultados» (ibíd.) y las circunstancias del profesor son diversas y no es probable que se puedan tratar de una forma rutinaria. Al contrario, «la buena práctica no puede equivaler a la *aplicación* directa de los conocimientos teóricos, sino que es una actividad que necesariamente implica el juicio y se basa en la experiencia tanto como en [...] el conocimiento científico» (ibíd.). En tercer lugar, duda de que sea útil considerar que la indagación del profesional es una investigación. Concede un lugar a las diferentes formas de inda-

gación práctica, tales como las investigaciones para la resolución de problemas y la reflexión. Pero, «aunque existen formas de indagación que están relacionadas estrechamente con las actividades prácticas y aunque éstas se puedan basar en ideas etnográficas, es importante distinguirlas de la etnografía» (Hammersley, 1992, pág. 152).

Tal vez siga habiendo rastros de elitismo académico en estas argumentaciones, pero los temas que plantean merecen una respuesta considerada desde el ámbito de la investigación del profesional. Dondequiera que se haya desarrollado, ha estado marcada con demasiada frecuencia por un pragmatismo ingenuo tal, que corre el peligro de quedar absorbida en la provisión de un apoyo técnico al *statu quo*. Sin embargo, caracterizar así todo el movimiento es rayar en lo caricaturesco.

LA EVALUACIÓN PARTICIPATIVA

Los métodos críticos de estudio de casos de Carr y Kemmis incluyen, como hemos visto, la participación de todos los interesados. Las formas cooperativas y participativas de evaluación se han desarrollado de diversas formas, aunque no es seguro aún hasta qué punto trascienden de la retórica. De momento son demasiado pocas las explicaciones sobre la evaluación participativa. Heron hace un resumen de los argumentos que se suelen aducir para este tipo de investigación. La indagación en la condición humana no es posible desde el exterior, sino que implica «nuestra propia acción personal e inteligente en cooperación con las demás personas con quienes decidimos compartir la condición dada» (Heron, 1996, pág. 201). Las personas llegan a ser tales sólo en relación con los demás. A menos que estemos en una relación mutua con ellas, no podemos investigarlas de una forma adecuada como personas: «Si no invitamos a participar a los sujetos humanos como coinvestigadores iguales, no están completamente presentes como personas y se convierten en personas sobornadas» (1996, pág. 202). Además,

[...] Investigar implica un compromiso educativo ineludible: facilitar en los sujetos de la investigación el desarrollo de su autodeterminación en la adquisición del conocimiento de la condición humana (1996, pág. 208).

Los investigadores participativos ponen mayor o menor énfasis en cinco compromisos (Cancian y Armstead, 1992):

1. Participación de las personas que se estudien.
2. Respeto por los conocimientos populares en el momento de acordar y realizar la investigación y al informar sobre ella.
3. Centrarse con talante liberador en los temas de la potenciación de las personas.
4. Una dimensión educativa mediante la concienciación.
5. La acción política y la transformación social.

Sin embargo, esto subestima la diversidad de la evaluación participativa. En el capítulo 4 distinguíamos las posiciones pluralista, reformista y emancipadora sobre la postura política de la evaluación. Todas ellas se pueden encontrar en la evaluación participativa, y todas se han adoptado para justificar los compromisos participativos. Los investigadores participativos tienen en común un grado de compromiso con una idea instrumental y activista de la investigación, pero algunos sostienen que la buena evaluación es transformadora porque es verdadera, mientras que otros dicen que es verdadera porque es transformadora. En primer lugar voy a ocuparme de esas formas de evaluación participativa que mantienen un compromiso más fuerte con la comprensión y el conocimiento, antes de considerar los planteamientos en que el cambio personal y político se considera como el principal, y quizás el único, criterio de validez.

La indagación participativa

Existen varias tendencias de indagación participativa, que reciben el nombre de Indagación Cooperativa, Ciencia de la Acción, Indagación de la Acción, e Investigación Participativa en la Acción, y están asociadas con la obra de Heron, Reason, Whyte, Schon, Argyris y Tarbert. En el apartado sobre la indagación reflexiva nos referiremos a la obra de Schon. De momento hablaré brevemente de la de Reason y de la de Heron. Aunque quizá sea una simplificación excesiva, los principales temas recurrentes de la indagación participativa son:

- La subjetividad: «La indagación humana válida esencialmente exige la total participación en la creación de conocimientos personales y sociales» (Reason, 1994b, pág. 332).
- Los modelos de la «ciencia normal» se sustituyen, en grado variable, por un énfasis en el valor del conocimiento profano y por el uso de unos métodos menos convencionales.

- El conocimiento surge *de* la acción y es *para* la acción. El objetivo es desarrollar una ciencia genuina de la acción, más que una ciencia aplicada: «El conocimiento práctico, el saber cómo, es la consumación, el cumplimiento, de la búsqueda del conocimiento» (Heron, 1996, pág. 34).
- El conocimiento se comprueba en los contextos de la acción directa.
- A diferencia de los enfoques más explícitos de la evaluación de defensa, el igualitarismo queda moderado por una confianza en el liderazgo.

Peter Reason ha formulado una de las argumentaciones más sólidas de la indagación participativa. Rechaza las visiones positivista y moderna del mundo alegando que separan al conocedor de lo conocido. Partiendo de la base de que «una metodología participativa» debe descansar en una visión del mundo participativa, está «muy convencido» de que «el propósito de la indagación humana no es tanto la búsqueda de la verdad como *curar* la alienación [...] que caracteriza a la experiencia humana» (Reason, 1994a, págs. 1, 10). Comparte con Heron la creencia en que «sólo podemos hacer una auténtica investigación *con* las personas si nos implicamos con ellas *como* personas, como sujetos y, por consiguiente, como coinvestigadores» (1994a, pág. 10).

Heron dice de la indagación cooperativa que implica a «dos o más personas que investigan un tema a través de la experiencia que de él tienen, utilizando una serie de ciclos en los que se mueven entre esta experiencia y la reflexión conjunta sobre ella» (Heron, 1996, pág. 1). La matización que hace de los tipos y niveles de participación se basa en lo que él describe como «realidad participativa» que da al paradigma participativo dos «alas»: la epistémica y la política. Por política entiende las decisiones referentes al razonamiento y a la toma de decisiones de la investigación. Por epistémica entiende las experiencias y la acción que son el objetivo de la indagación. Sobre esta base, distingue entre la participación política y la epistémica. Los temas clave son si, y en qué medida, los *investigadores* forman parte de la *experiencia* epistémica, y los *sujetos* forman parte de las *decisiones* políticas de la investigación. Presenta la naturaleza participativa de la indagación convencional (normalmente cuantitativa) tal como se muestra en la tabla 6.1. El contraste con la versión completa de la indagación cooperativa se muestra en la tabla 6.2.

TABLA 6.1. *Indagación convencional (típicamente cuantitativa).*

	Investigador	Sujeto
Participación en las decisiones	Completa	Ninguna
Participación en la experiencia	Ninguna	Completa

TABLA 6.2. *Indagación completamente cooperativa.*

	Investigador	Sujeto
Participación en las decisiones	Completa	Completa
Participación en la experiencia	Completa	Completa

Heron cree que la «investigación cualitativa [de corte tradicional] *sobre* las personas está a medio camino entre la investigación exclusiva y controladora *sobre* las personas, y la investigación completamente participativa *con* las personas» (1996, pág. 27). Los investigadores cualitativos comparten parcialmente la experiencia epistémica a través de métodos como el de la observación del participante y los sujetos comparten parcialmente las decisiones políticas a través de un consentimiento mejor informado. Pero el autor critica el énfasis que la investigación cualitativa pone en el acceso, como una «potenciación bajo los auspicios de una sutil opresión benigna» y «atrayendo a los guardianes y a los informantes para someterles al estudio con un diseño en el que no están invitados a colaborar y en el que, en el mejor de los casos, sólo están invitados a dar un consentimiento informado» (1996, pág. 28).

Heron admite la posibilidad de que la indagación cooperativa no alcance el ideal de la participación completa. Si, por un lado, se establece la separación entre lo epistémico y lo político, por otro, la indagación cooperativa no consigue separar «al inicio de la indagación, en la orientación metodológica facilitadora y del saber cómo, entre el investigador que toma la iniciativa y los coinvestigadores» (1996, pág. 23). El autor describe este hecho como «uno de los principales retos de la indagación cooperativa, y su talón de Aquiles extremadamente vulnerable» (1996, pág. 23). Además, es posible que la completa participación epistémica de los investigadores no sea viable, especialmente cuando son investigadores externos.

Participación en la evaluación de defensa

«Nosotros mismos somos reticentes a separar la epistemología de la ideología, las conclusiones de los deseos» (Stake, 1997, pág. 471). Stake plantea así el quehacer esencial de los enfoques de la evaluación en los que el cambio personal y político se consideran la principal referencia, y quizá la única, de la validez. Una vez más, la diversidad es el sello distintivo de este campo de la evaluación. Por ejemplo, la evaluación de defensa en los campos de la sanidad y el bienestar social casi siempre ha pretendido incluir a aquellos que se sitúan en el extremo de las prestaciones de los servicios sanitarios y de la atención social. Las versiones educativas de la evaluación de defensa han desarrollado con más frecuencia unas formas de participación que implican a otros profesionales más que a los niños y los alumnos.

Los primeros llamamientos a la evaluación participativa basaban su fuerza en una postura de *bienestar redistributivo*. Por ejemplo, Holman, el asistente comunitario e investigador de políticas sociales inglés, defiende una «investigación desde abajo» porque «los valores que defiendo me llevan a añorar el final de la pobreza y el fomento de la igualdad. Lo que me interesa de la investigación es simplemente esto: ¿de qué modo la investigación ayuda a los pobres?» (Holman, 1987, pág. 669). Critica gran parte de otros tipos de buena investigación acerca de la pobreza porque es *sobre* los pobres, y no se realiza *por* o *con* los pobres, y propone un modelo de abajo arriba en el que los investigados se convierten en los investigadores.

Se han realizado esfuerzos también para acercar la *investigación sobre el desarrollo del Tercer Mundo* a la práctica del desarrollo (Booth, 1994; Chambers, 1983; Edwards, 1989, 1994; Hall, 1979). Edwards parte de su experiencia de trabajo en organizaciones no gubernamentales cuando relaciona los problemas de

la «profesionalización» de los estudios sobre el desarrollo y la devaluación de los conocimientos populares; los valores y las actitudes de los investigadores y los profesionales que les impiden trabajar como iguales; el control del conocimiento por parte de las élites; y una incapacidad para unir la comprensión, la acción, la relevancia y la participación (1989, pág. 133).

El autor lamenta la creación y perpetuación de una serie de divorcios, el más completo de los cuales «está entre el producto de la investigación y los sujetos de esta investigación, los propios pobres». A

pesar de que insiste en que el método no ha sido el único problema, insiste también en que «la indagación participativa es el vehículo para comprender y cambiar el mundo simultáneamente» (1989, pág. 128). Según un certero proverbio que cita el autor, «se hace camino al andar».

La evaluación participativa de este tipo no es la panacea. Ha habido cierta exaltación generada por la colonización occidental de esta estrategia explícitamente liberadora, pero el entusiasmo se ha moderado en cierto grado por las dificultades de asegurar la auténtica participación en proyectos que casi siempre tienen su origen en «expertos». No se ha dejado de hablar de dos temas. En primer lugar, ¿a qué nos referimos cuando decimos que una evaluación es relevante? ¿Es la evaluación relevante lo mismo que la evaluación participativa? En un artículo posterior, parece que Edwards modera su posición. Acepta que no es necesario exigir que toda indagación sea participativa. Esto se puede aplicar de forma especial a la evaluación de los niveles más elevados de gestión de los programas de desarrollo. Sin embargo, «para ser relevante, la investigación *de algún modo* debe ir unida a la experiencia y las preocupaciones reales de las personas de a pie» (Edwards, 1994, pág. 284).

En segundo lugar, ¿cuál es la relación entre la evaluación de nivel básico y la evaluación de «nivel superior»? Uno de los peligros del cambio posmarxista hacia la adopción de explicaciones basadas en la cultura, la diversidad, la agencia y las construcciones del actor es que se ignora la relación entre el micronivel de la práctica individual y el macronivel. ¿Cómo se pueden comprometer con una evaluación básica quienes trabajan en el desarrollo y al mismo tiempo mantener con firmeza sobre la mesa los macrotemas?

Las estrategias participativas han derivado también de ciertas versiones de la *investigación feminista*. Por ejemplo, la teoría de la perspectiva feminista valora el igualitarismo epistemológico, se basa en una idea activista de la evaluación realizada con fines explícitamente políticos, está comprometida con la potenciación de las personas e implica un modelo de evaluación que tiende a la participación de los miembros de los grupos oprimidos en cada una de las fases de la actividad investigadora.

En la investigación y la práctica feministas ha habido una preocupación generalizada por el peligro de que la evaluación emancipadora conlleve unos riesgos de imposición (por ejemplo, Acker y otros, 1983; Ellsworth, 1989; Stacey, 1988). Veámos en el capítulo 5 la explicación de Whitmore de la evaluación de un programa prenatal pa-

ra mujeres solteras embarazadas, como ilustración de que se pueden conseguir unos resultados positivos. La evaluación la realizaron durante varios meses cuatro mujeres que habían seguido el programa, y Whitmore era la asesora del proyecto. En el grupo hubo tensiones y problemas de comunicación. Una de las mujeres lo abandonó y otra le decía a Whitmore:

Nuestro mundo es diferente del vuestro. El miedo se debe a que las personas siempre quieren humillarte, rebajarte [por estar en la asistencia social] [...] Nuestro modo de vida es distinto del vuestro. Simplemente, no confiamos en las personas como vosotros confiáis (Whitmore, 1994, pág. 92).

Pero la fuerza de lo que al final consiguieron demuestra que la evaluación participativa con grupos oprimidos no es sólo una retórica política.

La conclusión de Whitmore era que nunca pudo compartir por completo los mundos de las mujeres con las que trabajaba: «Mis mundos pequeños eran a menudo sus grandes mundos. Lo que yo suponía una “charla normal” para ellas eran “palabras de la profesora”» (1994, pág. 95). Martin (1994, 1996) llegaba a conclusiones parecidas en su propia investigación participativa feminista, y decía que tal investigación «pone unas expectativas irreales en la medida en que los investigados pueden intervenir en el proceso de la investigación» y que «incluso cuando los problemas constituyen una preocupación importante para las personas, [los investigados] tienen un trabajo y una vida privada que normalmente son prioritarios, mientras que la investigación ES el trabajo del investigador» (Martin, 1994, pág. 142).

Un ejemplo sensible y fascinante de investigación participativa es el que expone Miller, de tres años de investigación colaborativa con profesores que hasta ese momento habían sido sus alumnos de posgrado (Miller, 1990, 1992; Miller y Martens, 1990). Su finalidad es «extender el concepto de profesor como investigador a formas recíprocas e interactivas» con una agenda emancipadora (Miller, 1990, pág. ix). En la exposición se muestran los compromisos constantes del grupo mientras «siguen analizando de qué forma las intenciones y las formas de la pedagogía y la investigación emancipadoras nos atraen y nos confunden a la vez» (Miller y Martens, 1990, pág. 45).

Los grupos centrados pueden ser también un medio para la evaluación emancipadora. La explicación de Plaut del uso de los grupos centrados para la movilización de la comunidad entre las comunidades

rurales, políticamente conservadoras, de pobres y de blancos, y el trabajo de Jarrett con mujeres negras en peligro de pobreza a largo plazo son ejemplos de los beneficios potenciadores de las personas que se consiguen con los grupos centrados (Jarrett, 1993; Plaut y otros, 1993).

Stake tiene sus reservas respecto a la evaluación de defensa o propugnación: «Creo que nuestra postura debe ser la de oponernos a la defensa, minimizar la defensa, aunque siempre esté presente, ayudar a reconocer las defensas que tienen valor y las que no lo tienen, pero oponernos a compartirlas» (1997, pág. 474).

Los evaluadores comprometidos con la evaluación crítica pueden tener menos dudas respecto a las defensas buenas y malas. Pero todos los evaluadores participativos se enfrentan a la posibilidad de que su investigación se pueda utilizar para apoyar causas que desaprueben. Edwards no elude este problema y acepta la ambigüedad inherente de la evaluación participativa: «Por su propia naturaleza, el proceso de potenciación es incierto [...] su destino último es desconocido» (1994, pág. 288).

EVALUAR EN LA PRÁCTICA

Las pruebas de que se dispone indican que es poca o ninguna la relación entre el conocimiento relevante de la investigación y la capacidad de reflexionar críticamente sobre la propia práctica (Gibbs y otros, 1995). Los profesionales que entran en contacto con la investigación evaluativa se enfrentan a lo que se asemeja a un acertijo en lo que se refiere a la relación entre tal investigación y la práctica que ellos desarrollan. Una «solución» que antes señalábamos es aflojar el nudo gordiano de la práctica/investigación, afirmando para ello que la práctica es una especie de investigación, o que la investigación es una especie de práctica o servicio. Observábamos esta solución en algunas versiones de la investigación del profesional. Goldstein propugna una afinidad de este tipo cuando dice que «el lenguaje de la etnografía es el lenguaje de la práctica» y que «tanto el investigador cualitativo como el profesional dependen de talentos similares» (1994, págs. 46, 48).

En mi opinión, no es un argumento convincente. Los dilemas de la práctica y la evaluación no se resuelven con soluciones precipitadas, aunque muchas veces se diga lo contrario. Quienes abogan por una indagación y una evaluación reflexivas desarrollan una argumentación más sutil. La tesis subyacente en estos casos es que existe

una dimensión evaluativa en la práctica profesional, que es propia de la práctica, y no un «préstamo» de la metodología de investigación o evaluación. Veíamos en el capítulo 2 que Eisner ha desarrollado un modelo de este tipo de indagación reflexiva para la educación, basado en la especialidad del entendido y en la crítica y en el que la indagación se considera más semejante al arte que a las interpretaciones tradicionales de la ciencia. Schon ha desarrollado un modelo de indagación más amplio, que pretende abordar el carácter de trabajo reflexivo de todas las profesiones (Schon, 1983, 1992, 1995; para un enfoque comparable para la asistencia social, véase, por ejemplo, England, 1986).

Eisner y Schon coinciden en rechazar las antiguas concepciones positivistas del conocimiento y de la relación de éste con la teoría y la acción. Por ejemplo, Eisner (1988) critica los planteamientos científicos y tecnológicos de la escuela por cuatro razones. Primera, «la exclusividad de lo particular se considera un “ruido” en la búsqueda de las tendencias generales y de los efectos principales» (1988, pág. 139). Segunda, la orientación tecnológica de la práctica «tiende a fomentar un objetivo primario en la consecución de algún estado futuro y, de paso, tiende a socavar la importancia del presente» (1988, pág. 140). Los objetivos siempre son inalcanzables. Tercera, se pretende objetivar el conocimiento, de modo que las calidades que uno decide considerar deben ser manifiestamente empíricas. Por último, los test estandarizados de las escuelas actúan como si ellos mismos fueran objetivos educativos, y la uniformidad se convierte en una aspiración. Las decisiones adoptadas sobre las formas de evaluación tienen consecuencias para el carácter de la enseñanza, el currículo y las metas educativas. Schon extiende esta crítica de la racionalidad técnica al modelo tácito de relación entre investigación y práctica:

Se presume que la investigación y la práctica están unidas por un intercambio en el que los investigadores ofrecen unas teorías y unas técnicas aplicables a los problemas de la práctica, y los profesionales, a su vez, presentan a los investigadores nuevos problemas en los que trabajar y unos test prácticos de la utilidad de los resultados de la investigación (Schon, 1992, pág. 53).

Las pruebas empíricas

En recientes trabajos empíricos se ha visto que los profesionales pretenden comprender y resolver los temas evaluativos en su trabajo cotidiano. Elks y Kirkhart dicen que «se necesita un modelo de investigación alternativo, un modelo que sea explorador más que confirmador, que construya un modelo de evaluación a partir de las propias explicaciones del profesional, y no que imponga un ideal para comprobar si uno se ajusta a él» (1993, pág. 555). Los autores entrevistaron a diecisiete trabajadores sociales y les preguntaron cómo evaluaban lo que hacían, y cómo sabían que hacían las cosas bien o mal. Los profesionales reconocieron la dificultad de saber si eran eficientes. También percibían la existencia de una incompatibilidad entre los roles de evaluador y profesional. Los investigadores señalan que los profesionales adoptan un modelo implícito de práctica de la evaluación que ellos describen como «modelo pragmático profesional». Tal modelo incluía una confianza en la intuición y la experiencia, una idea interiorizada de profesional ideal, una dependencia de la retroalimentación de los colegas, los amigos y la familia, y un modelo de cliente ideal que siempre incluía el crecimiento y el cambio.

Humphrey y Pease realizaron un estudio en el que entrevistaron a funcionarios ingleses encargados del seguimiento de las personas que están en libertad condicional, para preguntarles lo que pensaban de la efectividad de ésta. Los funcionarios tendían a «disociar» el proceso de supervisión de los resultados de la libertad condicional. Así, una persona decía que los clientes «a menudo tienen más problemas, pero no creo que en modo alguno sea un reflejo de si yo he sido o no eficiente». En efecto, existía la creencia generalizada de que en el hecho de ser eficiente intervenía la suerte. Uno podía hacer un trabajo «brillante», pero si las circunstancias le eran adversas resultaba que seguía haciendo mal las cosas. De modo que «si se considera que la suerte determina el resultado, la supervisión de la libertad condicional se convierte en una simple cuestión de mantener a un incompetente en la comunidad esperando que la suerte le sea propicia» (Humphrey y Pease, 1992, pág. 40).

Los trabajos posteriores de Shaw y Shaw indican que los asistentes sociales parecen tener en mente dos modelos opuestos de evaluación: una «evaluación como Dios manda» formal y la autoevaluación. La evaluación formal se experimenta como muy ajena a las realidades de la asistencia social y en contraste casi completo con los «mapas» evaluativos que los trabajadores sociales hacen de su evaluación

cotidiana real (Shaw y Shaw, 1997b). Las estrategias de evaluación se construían a partir de un «plan de juego», cuyo éxito, al igual que en la investigación de Humphrey y Pease, se consideraba en parte contingente en función de la intervención adversa de la «mala suerte». Para juzgar su práctica, los asistentes sociales consideraban si su trabajo producía unas recompensas emocionales; si el caso era «emotivo»; si la intervención producía un cambio directo e incremental; si la práctica se desarrollaba sin que el funcionamiento del sistema de bienestar causara perjuicio alguno; y si se disponía de pruebas positivas por parte de otros colegas. A esos profesionales les preocupaban las causas y las razones de los resultados de su trabajo, tenían unas ideas muy elaboradas sobre la complejidad y la ambigüedad de las pruebas de la asistencia social y eran conscientes de la constante interacción del conocimiento y del sentimiento en la práctica (Shaw y Shaw, 1997a). En esta importancia de las pruebas resuena la observación de Erikson de que «las pruebas no están completas si el médico no consigue utilizar sus propias reacciones emocionales durante un encuentro clínico, como fuente de pruebas y como guía para la acción» (Erikson, 1959, pág. 93).

La indagación reflexiva

No hay duda de que Donald Schon comentaría las reflexiones de estos profesionales a la luz de cómo entiende él la indagación reflexiva. Para este autor, nuestro conocimiento es algo generalmente tácito y está *en* nuestra acción y, por consiguiente, nuestra vida de todos los días revela «un patrón de conocimiento en acción tácito» (Schon, 1992, pág. 56).

En su práctica diaria, hace innumerables juicios de calidad para los que no puede aportar unos criterios adecuados, y muestra unas habilidades para las que no puede aportar las normas y los procedimientos (Schon, 1983, pág. 50).

En la introducción de este capítulo nos referíamos al conocimiento tácito de este tipo. Pero el conocimiento tácito se desvanece cuando nos encontramos con sorpresas debidas a que nuestras reacciones espontáneas ante los fenómenos de la vida cotidiana no siempre funcionan. Esto produce a menudo lo que Schon denomina la «reflexión en la acción», por la que los problemas se resuelven mediante la «in-

vención en línea» y la experimentación *in situ*. «Estimulados por la sorpresa, piensan de nuevo en la acción y en el conocimiento que está implícito en la acción» (Schon, 1983, pág. 50).

La exposición que Schon hace de la reflexión en la acción reúne las siguientes cualidades de la evaluación reflexiva:

- Es una actividad disciplinada en la que las fuentes de la comprensión y la acción no se eliminan automáticamente de la superficie: «La acción hábil suele reflejar un conocimiento mayor del que imaginábamos» (1983, pág. 51).
- La evaluación reflexiva implica una reflexividad, una comprobación mediante la reflexión. Es oportuno aquí el mandamiento de Erikson. Debemos intentar constantemente «forzar nuestros supuestos favoritos para que se conviertan en inferencias probables» (Erikson, 1959, pág. 94). Los profesionales no son simples usuarios del conocimiento, sino sus generadores.
- Existe una posibilidad inherente de comparar las capacidades «profesionales» y las de las personas profanas para mostrar los elementos de la indagación reflexiva.

Schon perfila aún mejor esta imagen cuando describe la relación que entiende que existe entre el profesional y el cliente. La relación

adopta la forma de una conversación literalmente reflexiva [...] Atribuye a su cliente, como a sí mismo, la capacidad de querer decir algo, de saber y de planificar. Reconoce que sus significados pueden tener para su cliente un sentido distinto del que pretende [...] Reconoce la obligación de hacer accesibles sus propias interpretaciones a su cliente [...] Intenta descubrir los límites de su experiencia a través de la conversación reflexiva con el cliente (1983, págs. 295-296).

Se pueden hacer varias críticas a la obra de Schon. No reconoce los temas de poder ni en la relación ni en el servicio. Tampoco parece que contemple ninguna dimensión directamente participativa. Sus ejemplos se centran mucho en el profesional que habla con los colegas y los supervisores, sin que intervengan los alumnos, los niños o los usuarios de otros servicios (Schon, 1995). Es posible que subestime también el nivel de decisiones que comporta la práctica diaria en los hospitales, en las escuelas y en el trabajo social (Bloor, 1978). Sin embargo, su análisis es original y hace reflexionar. Es el único teórico del campo de la evaluación reflexiva y del profesional que ha de-

sarrollado un modelo que no es específico de ninguna profesión. Esto supone en sí mismo una agenda para la evaluación etnográfica comparativa de los servicios que aún no se ha abordado. Las preocupaciones de Schon son relevantes para la evaluación cualitativa, aunque no la ejemplifiquen de forma activa. No obstante, existe toda una literatura nueva sobre la indagación cualitativa reflexiva (Fook, 1996; Gould, 1999; Gould y Taylor, 1996).

LA EVALUACIÓN CUALITATIVA PARA LOS PROFESIONALES

En este capítulo nos hemos centrado en la evaluación del profesional «desde dentro». No quisiera que se me malinterpretara en este punto. Algunas veces han dicho los evaluadores que la evaluación desde dentro es más válida que la que se realiza desde fuera, alegando que los profesionales tienen acceso a sus propias interpretaciones, sus motivos y sentimientos; que conocen la situación de primera mano; y que están en mejores condiciones de poder acceder a las cosas, etc. Mi insistencia en la evaluación interna no refleja esta idea. Estoy de acuerdo con Hammersley en que ninguna posición garantiza un conocimiento válido, como ninguna lo impide tampoco (Hammersley, 1993a).

Sin embargo, y en consecuencia, debemos concluir que la evaluación del profesional puede partir también de la investigación evaluativa que tiene una relevancia más o menos directa con la práctica. No me refiero a las etnografías como la investigación de Atkinson (1995) sobre la charla médica y el trabajo médico, el estudio de Pithouse (1998) sobre un equipo de atención infantil de un Departamento de Servicios Sociales inglés, ni los estudios interaccionistas de Hargreaves sobre la conducta en el aula. Son estudios que tienen una relevancia evaluativa, pero sus implicaciones para la práctica directa siguen siendo implícitas.

No obstante, la atención que se ha prestado a los *usos evaluativos de la etnografía* ha sido insuficiente. Si recordamos las sonoras quejas de McNamara (1980) sobre la irrelevancia de la etnografía, tendremos en que se necesita trabajar más en los usos de la etnografía por parte del profesional. En este sentido, Bloor y McKeganey y Silverman constituyen unos precedentes que no suelen ser habituales entre los etnógrafos (Bloor, 1997a; Bloor y McKeganey, 1987, 1989; Silverman, 1993, 1997a). Bloor no es optimista sobre la posibilidad de que la investigación tenga un impacto político. A diferencia de

Weiss y otros autores, que se centran casi exclusivamente en los usos de la evaluación para el proceso político, Bloor sostiene que «las verdaderas oportunidades de influencia sociológica se sitúan más en el tajo que en el despacho del jefe, que las verdaderas oportunidades de influencia sociológica están en las relaciones con los profesionales, no con quienes dirigen la práctica» (1997a, pág. 234).

En estudios de etnografía sobre la conducta de riesgo relacionada con el sida en la prostitución masculina de Glasgow, y en otros estudios etnográficos comparativos de ocho comunidades terapéuticas, el autor señala que la etnografía puede serle útil al profesional de dos formas. En primer lugar, puede «modelar» una prestación de servicios que se puede transferir a los suministradores del servicio. Por ejemplo, «el trabajo de campo etnográfico, en sus contactos regulares y prolongados con los sujetos de la investigación, tiene mucho en común con el trabajador de los servicios sociales que promueve que se pidan las prestaciones a las que se tiene derecho» (Bloor, 1997a, pág. 227). Este punto tiene ciertas similitudes con la idea de los evaluadores reflexivos de que la etnografía se puede colonizar y traducir a formas que se puedan emplear en la práctica: «Una buena descripción de determinados tipos de práctica terapéutica puede ayudar a los profesionales a emitir juicios evaluativos sobre su propia práctica» (1997a, pág. 229). En segundo lugar, los etnógrafos pueden, donde corresponda, dirigir la atención de los profesionales hacia prácticas que consideren que merece la pena difundir y prestar una mayor consideración. Bloor y McKeganey (1989) hacen una relación de siete prácticas que en su opinión fomentan la terapia en sus enclaves originales, y que los autores debatieron con los profesionales de distintas comunidades terapéuticas.

Señalan los autores las correspondientes implicaciones para los etnógrafos, en el sentido de que «cualquier intento de explotar aún más el potencial evaluativo de la etnografía para un público profesional debe ir paralelo con un avance de los estudios etnográficos que se centren en el trabajo de los profesionales, no en su conducta» (Bloor y McKeganey, 1989, pág. 210). Hasta que llegue ese momento, cuando «los propios ciudadanos encargan el trabajo a los profesionales, el lugar de los sociólogos no está en lamentar una falsa conciencia ni en exigir una oposición a los servicios de bienestar social» (Bloor, 1997a, pág. 235).

En los dos últimos capítulos hemos analizado las consideraciones especiales que se plantean en la realización de la evaluación cualita-

tiva de programas, de políticas y de la práctica individual. Las diferencias entre ellas están lejos de ser superficiales o marginales. Pero estas diferencias no constituyen una razón para fragmentar la investigación evaluativa y de políticas en múltiples formas diferenciadas de indagación, cada una con sus propios axiomas y su metodología. En los capítulos siguientes pretendemos explicar que los propósitos evaluativos definen enfoques diferenciados pero coherentes de las decisiones adoptadas en relación con el diseño, el trabajo de campo y el análisis de evaluación cualitativa.

CAPÍTULO 7

LAS DECISIONES SOBRE EL DISEÑO DE LA EVALUACIÓN

El etnógrafo «inscribe» el discurso social; lo anota. Al hacerlo, lo convierte de un acontecimiento pasajero, que sólo existe en su propio momento de ocurrencia, en una explicación, que existe en sus inscripciones y se puede reconsultar.

GEERTZ (1973, pág. 20).

La idea que Geertz tiene sobre las características de las descripciones etnográficas, en el sentido de que describen el flujo del discurso social y tratan de «rescatar lo “dicho” de tal discurso de sus ocasiones perecederas y fijarlo en términos que se puedan leer detenidamente» (1973, pág. 20), refleja los aspectos clave de la concepción de la metodología de la evaluación cualitativa que describimos en los tres capítulos siguientes.

Las divisiones de los capítulos son convencionales, al menos en un sentido superficial. Al diseño de la evaluación le sigue la recogida de datos en el capítulo 8 y el análisis de la evaluación en el capítulo 9. Pero en la práctica de la evaluación cualitativa puede ocurrir que normalmente no sea posible esta distinción tan definida de tres fases en la búsqueda del conocimiento. Desde luego es posible que no existan como operaciones autónomas. Dedicamos un espacio importante de este capítulo a la ética de la evaluación, a la evaluación participativa y al trabajo cualitativo sobre los resultados de las políticas, los programas y las prácticas. Sin embargo, éstos no son más exclusivamente temas de «diseño» que temas del proceso del trabajo de campo o del análisis, y todos ellos reaparecen en los capítulos siguientes.

Aparecen dos cuestiones preliminares. ¿Cómo incidirá la evaluación cualitativa en las decisiones sobre el diseño? ¿Existen unos métodos específicos de evaluación cualitativa? La postura que hemos adoptado a lo largo de este libro es que no existe una distinción me-

metodológica intrínseca entre investigación cualitativa y evaluación cualitativa. Pese a los argumentos en contra de autores como Lincoln y Guba (1986b), no veo contradicción alguna en la expresión «investigación de evaluación». La idea asumida en este libro es que existe un grupo de propósitos de evaluación reconocibles que se solapan en parte, pero no por completo, con los propósitos de la investigación. Así pues, las decisiones de diseño de la evaluación y la investigación parten de las mismas diversas opciones de métodos. En lo que se distingue la evaluación es en la dirección que la evaluación de las políticas, los programas o las prácticas nuevas o establecidas da a la elección de los métodos y a los usos que de ellos se hace: «El diseño es la secuencia lógica que conecta los datos empíricos con las preguntas de investigación iniciales de un estudio y, en última instancia, con sus conclusiones» (Yin, 1994, pág. 19). En este sentido, «el diseño de una investigación supone un problema *lógico*, no un problema *logístico*» (1994, pág. 20). La lógica y los valores de la evaluación configuran las decisiones referentes al diseño. Desde el punto de vista logístico, la evaluación cualitativa difiere poco de la investigación cualitativa. De modo que el diseño de la evaluación «se centra en formular las preguntas correctas y en seleccionar los métodos apropiados, para que se puedan hacer afirmaciones creíbles sobre el fenómeno que se estudie» (Fournier y Smith, 1993, págs. 315-316).

Sin embargo, me aparto claramente de la postura general referente a las decisiones sobre el diseño de la evaluación, la postura que definíamos de «caballo según carrera». La idea rectora es que los diseños cualitativos, que muchas veces se identifican con los enfoques del estudio de casos, son más adecuados para responder a las preguntas sobre los *procesos*. Para las preguntas sobre los *resultados*, se dice, sirven mejor los diseños comparativos tradicionales que se ocupan de medir los resultados. En esta distinción, a veces hay algo más que un simple indicio de menosprecio de la evaluación cualitativa, como lo ilustra, por ejemplo, el comentario de Scriven de que la evaluación formativa en los procesos y la mejora de los programas es, en realidad, una versión provisional de la evaluación sumativa, que facilita información sobre el «valor medio» (1997, pág. 499). Shadish y sus colegas son menos tendenciosos que Scriven, pero el efecto global es similar (Shadish y otros, 1990). Su conclusión es que los diseños cualitativos tienen sentido cuando:

- El evaluador desea disponer de amplitud.
- Se conocen pocas preguntas antes de la evaluación.

- La evaluación la van a utilizar unos lectores que no pueden experimentar el programa por sí mismos.
- Los evaluadores pueden renunciar a respuestas de mayor calidad ante preguntas más específicas.
- No tendrá prioridad un sumario sucinto de los resultados.
- Las generalizaciones entre diversos enclaves no resultan prioritarias.
- El descubrimiento tiene mayor prioridad que la confirmación.
- El cliente de la evaluación considera que la evaluación cualitativa es creíble.

Concluyen los autores que los estudios cualitativos de casos pueden ser útiles en las primeras fases de la puesta en marcha de un programa nuevo. En otras palabras, hacen un refrito de la idea ya tan desprestigiada de que la investigación cualitativa desempeña una función exploratoria previa a la comprobación. Nos queda una imagen de la evaluación cualitativa como un ejercicio impreciso, de objetivos vagos, descriptivo e inductivo, sólido en lo que a la experiencia indirecta se refiere, pero crónicamente en peligro de perder la credibilidad a los ojos de las personas que importan: los clientes.

Rechazo esta idea de combinar las alternativas cualitativa y cuantitativa. La evaluación cualitativa problematiza y rechaza los estereotipos convencionales sobre el alcance respectivo y la división del trabajo de ambas. Por ejemplo, veíamos en el capítulo 1 que Williams hace un inventario más imparcial y que hace pensar en preguntas diferenciadoras en respuesta a la cuestión: «¿Cuándo es apropiada la evaluación naturalista?».

Consideremos los siguientes ejemplos de su lista:

- ¿Las definiciones oficiales del evaluando son suficientes o se requieren unas perspectivas émicas?
- ¿Es necesaria la evaluación de los procesos con los que el evaluando plantea sus resultados?
- «Si el resultado incluye las acciones complejas de enclaves naturales [...] un componente de naturalismo facilitaría [...] unos juicios más fiables y creíbles» (Williams, 1986, pág. 88).
- ¿Hay tiempo para estudiar al evaluando a través de su ciclo natural?
- «¿Se puede estudiar al evaluando sin molestar, tal como opera, de forma ética?» (1986, pág. 89).

Estas preguntas implican una lista de criterios muy distintos de los que ofrecen los teóricos de la evaluación de tipo tradicional. A lo largo de este capítulo decidimos el orden de prioridad de los temas referentes a la ética y, al hacerlo, sostenemos que las preocupaciones sobre la justicia social deben ocupar un lugar prioritario en la evaluación cualitativa. Rechazamos el supuesto de que la evaluación cualitativa se frustra por las cuestiones referentes a los resultados y defendemos que los estudios de casos y otros diseños cualitativos pueden permitir una comprensión de las causas y los resultados que en muchos casos se ajusta mejor a los propósitos de la evaluación que otros diseños más convencionales. Recuperamos el debate sobre la evaluación participativa y reflexiva que ocupaba el capítulo 6, y terminamos el capítulo considerando cómo configuran y enriquecen el diseño de la evaluación. Antes de ocuparnos de los temas de la ética y de la justicia social, consideramos brevemente el problema preliminar de si una evaluación es viable en un caso determinado.

VALORAR LA EVALUABILIDAD

Wholey delimita varios problemas que, si no se resuelven, harán que cualquier evaluación resulte insignificante. Dice que:

- La falta de acuerdo sobre los objetivos, los efectos secundarios y los criterios de aplicación se traducirán en una evaluación que se centre en cuestiones no representativas o irrelevantes.
- Si los objetivos y las metas no son realistas, teniendo en cuenta el nivel de los recursos y las actividades del programa, éste «fracasará» inevitablemente.
- Cuando no se pueda proporcionar una información relevante sobre el programa, la evaluación no será concluyente.
- Si los responsables políticos o los gestores no están dispuestos o no pueden actuar, partiendo de la evaluación, para cambiar el programa, entonces el resultado será una «información en busca de usuario» (Wholey, 1994, pág. 16).

La valoración de la evaluabilidad muestra si un programa se puede evaluar de forma significativa y también si es previsible que se traduzca en acción para mantener o cambiar la política, el programa o la práctica que se evalúen, o algún futuro evaluando.

El análisis que Wholey hace de los problemas y de sus consecuencias es, por lo general, útil. Subraya los factores que evitan una aplicación no problemática del modelo de evaluación de objetivos y resultados y deja claro que existen unas actividades relacionadas con la evaluación que posibilitan una mejor comprensión y una mejora del proceso evaluador. En su formulación quedan algunos problemas. Por ejemplo, da un relieve exagerado a las metas y los objetivos, y su análisis refleja su idea de que los gestores del programa son los principales destinatarios de la evaluación. Pero la dificultad más grave está en la solución que propone. Su solución es que hay que aclarar los diseños del programa, hay que analizar las realidades del programa, y hay que rediseñar los programas cuando sea necesario, de manera que:

- Estén bien definidas las metas, los objetivos, los efectos secundarios y las necesidades de información prioritaria.
- Se cambien los recursos, las actividades y las metas para que sean verosímiles.
- Estén presentes las medidas relevantes y viables para la aplicación del programa.
- Exista un acuerdo entre los futuros usuarios sobre cómo se va a utilizar la información.

Algunos de los problemas de la lista inicial de Wholey *se pueden* atenuar, pero la evaluación no puede estar supeditada a su solución mediante la fórmula racional que aquí se indica. Siempre estarán presentes en mayor o menor grado. Por ejemplo, la distinción que el autor hace entre efectos buscados y secundarios delata una imagen mecánica o quizá química de los programas. Wholey relega también la política de la evaluación a un estatus invisible y no problemático. Por ejemplo, no piensa en que el acuerdo en centrarse en las metas del programa supone excluir los problemas que Weiss y otros autores plantean respecto a la legitimación tácita de las metas del *establishment* por parte de los evaluadores (House, 1993; Weiss, 1987). Su idea de que se puede alcanzar un acuerdo previo sobre cómo se va a emplear la información de la evaluación deja que se cuele el supuesto implícito de que el uso de la evaluación es principalmente instrumental. Hemos visto en los capítulos 4 y 5 que se trata, en el mejor de los casos, de una idea tendenciosa del uso de la información y, la mayoría de las veces, de una idea excesivamente racionalista. Por último, su idea de la valoración de la evaluabilidad previa no cabe en

modo alguno en ninguna práctica participativa. Aunque habla de implicar a los futuros usuarios de la evaluación y de no imponer los supuestos del experto, las personas en las que piensa son las responsables de los presupuestos, los gestores de mayor responsabilidad y el personal de mayor autoridad en el proyecto, como veíamos en la exposición que hacía del Programa Prenatal de Tennessee dentro del capítulo 5.

La valoración de la evaluabilidad puede ser o no ser apropiada. Presupone la posibilidad de una estrategia consensual para la evaluación, y como tal quizá se ajusta mejor al modelo de evaluación cualitativa que se basa en una postura receptiva hacia los interesados.

EJEMPLO 7.1. LA VALORACIÓN DE LA EVALUABILIDAD

Pulice (1994) ilustra una valoración de la evaluabilidad centrada en el interesado al hablar de los métodos de evaluación cualitativa en el sector público de la atención a personas con trastornos mentales graves. Recomienda identificar los grupos relevantes de clientes a todos los niveles y completar las entrevistas preliminares con los informantes clave sobre la razón de la solicitud de la evaluación, el marco temporal probable y sus ideas respecto a los usos futuros de la evaluación. Asimismo aconseja que se haga una recopilación de legislación, de documentos relevantes y de informes de comisiones «para determinar la naturaleza y el contenido de la documentación relacionada con la pregunta de la evaluación» (1994, pág. 307).

Luego sugiere que el evaluador o los evaluadores desarrollen una taxonomía preliminar de los grupos de clientes identificados desde el punto de vista de su previsible papel como defensores, planificadores, facilitadores, controladores o gestores. Por último, recomienda realizar unas entrevistas semiestructuradas que se centren en las grandes preguntas de la evaluación. En esta fase sería posible usar grupos centrados.

Las valoraciones de la evaluabilidad reflejarán las perspectivas subyacentes de los evaluadores. Por ejemplo, los evaluadores críticos querrán explorar, con el ánimo de moderarlas, las posibilidades de imponer sus propias visiones del mundo a quienes participan en la evaluación. Cuando tal imposición parezca inevitable, se puede decidir que la evaluación no es viable. Los evaluadores que parten de una postura constructivista y relativista querrán analizar la viabilidad de una auténtica evaluación que tenga en cuenta el contexto local y po-

tencie a los interesados. Rodwell y Woody piensan que «los métodos constructivistas para determinar la evaluabilidad son muy apropiados cuando el objetivo es implicar en el proceso a todos los grupos de interesados que sea posible, al tiempo que los problemas se plantean a partir de la observación, de las experiencias y de los datos» (1994, pág. 319).

Describen estos autores una propuesta de evaluación —que se malogró— de un entorno clínico: «Ignorábamos el contexto de los hechos» (1994, pág. 324). Cuando, después de valorar la evaluabilidad, se hizo la propuesta, se desestimó, alegando que no seguía la política de investigación de la agencia en cuestión; que amenazaba la homeostasis de la organización debido a la amplitud exagerada del objetivo y a que había poco tiempo para alcanzarlo; y que no se necesitaba un estudio detallado de ese tipo, sino que el organismo quería unos estudios comparativos de servicios similares en los que se utilizaran métodos cuantitativos. Los autores piensan que su fracaso se debió en parte a que no tuvieron en cuenta los criterios de autenticidad. Tales valoraciones «crean un referente en un entorno de programa de evaluación no muy racional» (1994, pág. 326).

ÉTICA DEL DISEÑO Y JUSTICIA SOCIAL

Con frecuencia, los problemas éticos de la evaluación se han abordado como si fueran más o menos sinónimos de los problemas éticos de la experimentación o la cuasi-experimentación. Afortunadamente, la experiencia y la reflexión de los investigadores y los evaluadores cualitativos nos han rescatado de esa miope perspectiva sobre la ética de la evaluación. House (1993) distingue cuatro situaciones en las que el evaluador se enfrenta a problemas éticos. Cuando:

1. Se oculta la naturaleza de la investigación de evaluación a los participantes.
2. Se expone a los participantes a actos que les pueden perjudicar.
3. Se invade la intimidad de los participantes.
4. Se ocultan los beneficios a los participantes.

Pensemos, por ejemplo, en el primero de estos problemas. La evaluación encubierta es un «ejemplo particularmente peligroso de investigación social encubierta, porque las conclusiones no se usan tan sólo para iluminar nuestros conocimientos del mundo social, sino

posiblemente para cambiarlo» (Finch, 1986, pág. 203). Las formas de evaluación de abajo arriba corren un menor peligro de encontrarse con este problema, debido a la cercanía de los implicados y al tipo de participación que se da en esta clase de evaluación. Sin embargo, sigue habiendo lo que Stacey (1988) denomina la ilusión de alianza y el problema de que los participantes revelen más de lo que pretenden. En la evaluación tradicional, es probable que el participante tenga relativamente claro cuándo el evaluador «está trabajando» y cuándo «descansa». Lo previsible es que los participantes supongan que cuando están en situaciones informales, el evaluador está descansando. Pero no es probable que ocurra lo mismo en la evaluación cualitativa, de modo que los participantes corren el riesgo de desvelar cosas que no quisieran.

Se ha apelado muchas veces al principio del consentimiento informado, para proteger de tal abuso a los participantes. El problema se plantea, por supuesto, porque esto «implica que el investigador sabe *antes* de que suceda algo [...] qué es lo que va a suceder y cuáles serán sus posibles efectos» (Eisner, 1991a, pág. 214). No suele ocurrir así, especialmente en la evaluación cualitativa.

EJEMPLO 7.2. EL CONSENTIMIENTO INFORMADO

Al personal de una residencia de ancianos se le dice:

Pedimos su permiso para observar su residencia durante ocho semanas. Al final de las ocho semanas, si ustedes lo desean, gustosamente les facilitaremos una retroalimentación sobre lo que hayamos visto y sobre lo que a nuestro parecer significa.

Este planteamiento tiene la virtud de destacar el carácter voluntario del acuerdo. Explica la duración y la forma general del estudio y promete una retroalimentación para quienes la soliciten. Pero hay algunas preguntas que no se responden:

- ¿Se publicará el estudio?
- ¿Para qué lectores?
- ¿El personal tendrá derecho a leer y comentar el informe antes de que se publique?
- ¿Se incluirán sus comentarios en el informe si creen que ha habido interpretaciones y/o representaciones muy equivocadas?
- ¿Se entrevistará a los residentes?

- De ser así, ¿cómo se negociará *su* consentimiento informado?
- ¿Cuál será el papel de los investigadores durante el período de observación?
- ¿La evaluación supondrá un trabajo adicional para el personal de la residencia?
- ¿Cómo se protegerá la confidencialidad?

Algunos participantes serán menos sutiles que otros al plantear cuestiones:

¿Los investigadores tienen la responsabilidad ética de desempeñar un doble papel: primero, el de investigadores con un proyecto destinado a satisfacer los propósitos de su investigación y, segundo, el de defensores de los profesionales, planteando preguntas que los investigadores saben que hay que plantear para que los profesionales valoren adecuadamente los riesgos? (Eisner, 1991a, pág. 217).

Los problemas de la *confidencialidad* son también más agudos en la evaluación cualitativa. Los evaluadores cuantitativos muchas veces pueden abordar los temas de la confidencialidad mediante un proceso de muestreo y con unas garantías técnicas cuando se analizan los datos. Podemos decidir garantizar el anonimato de las revelaciones. ¿Pero tienen el mismo derecho a la intimidad todos los participantes en una evaluación? Las revelaciones que en la prensa amarilla se hacen sobre la vida de los famosos a menudo se tratan de forma distinta a como se tratan las revelaciones sobre personas más humildes. ¿El compromiso con la intimidad abarca todas las situaciones? Es posible que tengamos muy claro qué haríamos si viéramos que un profesor acaricia con ánimo lúbrico a un niño. Pero ¿y en el caso de incompetencia flagrante de un profesional? ¿O ante incidentes que consideramos ejemplos manifiestos de racismo? La respuesta no está tan clara. Es posible que la escuela o la entidad ya conozcan la situación. Además, delatar unos hechos durante la evaluación puede impedir la posibilidad de delatarlos en el informe final. Y, aunque se plantee con menor frecuencia, ¿qué debería hacer el evaluador si se encuentra con que una persona a quien las autoridades consideran deficiente o «difícil» en su trabajo, demuestra unos logros destacables y que no se le reconocen? La cuestión es prácticamente la misma que la que se planteaba respecto al consentimiento: «¿Los investigadores tienen la responsabilidad ética de promover el trato justo de aquellos a quienes observan?» (Eisner, 1991a, pág. 219).

En Estados Unidos se han desarrollado los criterios de la evaluación (Stufflebeam, 1991), pero no se ha resuelto el problema de aplicar los principios éticos a la realización de la evaluación. House piensa que el equilibrio de los principios en las situaciones concretas es «el acto ético definitivo» (1993, pág. 168). Por ejemplo, es posible que la adhesión a los principios éticos demuestre ir en detrimento de la eficacia del programa (Lewis, 1988). El director de un hospital puede decidir disminuir el volumen de trabajo para aligerar la carga del personal y, por lo tanto, reducir el número de pacientes que se pueden visitar. Al reflexionar sobre su propia experiencia de errores éticos, Eisner concluye:

Quizá queramos asegurarnos un consentimiento informado, pero sabemos que no siempre podemos informar porque no siempre sabemos las cosas. Nos gustaría proteger la intimidad personal y garantizar la confidencialidad, pero sabemos que no siempre podemos cumplir tales garantías. Nos gustaría ser sinceros, pero a veces la sinceridad es inapropiada. No nos gusta pensar que utilizamos a los demás como recurso para alcanzar nuestros propios fines profesionales, pero si nos embarcamos en un estudio de investigación que nosotros conceptualizamos, dirigimos y redactamos, es prácticamente seguro que utilizaremos a los demás para nuestros propósitos (Eisner, 1991a, págs. 225-226).

Evaluar para la justicia social

Todo lo referente a la justicia social y la evaluación se ha tratado de forma incompleta, aunque cada vez es más abundante la literatura sobre el tema. Los evaluadores «no viven en un estado de gracia metodológica» (House, 1991a, pág. 245) y las consideraciones referentes a la justicia inciden en las decisiones adoptadas sobre el diseño de la evaluación. Me ocuparé brevemente de dos enfoques, para señalar de forma sucinta las ideas clave, poner un ejemplo e indicar las cuestiones que cada uno de ellos plantea.

La justicia como imparcialidad. Uno de los ejemplos más claros de esta postura es el de Ernest House. Para él, ninguna de las teorías dominantes de la justicia es completamente satisfactoria como base para la evaluación. Aboga por una base moral de la evaluación que se asiente en los principios de la igualdad moral, la autonomía moral, la imparcialidad y la reciprocidad, sin estar seguro de cómo hay que equilibrar mutuamente todos estos principios en cada situación. Ninguno

debe tener prioridad y las decisiones se deben tomar de forma pluralista, de modo que, aunque las consideraciones sobre la eficacia cumplan su parte, la justicia debe ir por delante. En otras palabras, House se aproxima mucho a John Rawls y también a planteamientos más intuicionistas, con una pequeña dosis de eficacia utilitaria.

Pone el ejemplo de negociar un acuerdo de evaluación justo y exigente (House, 1980, cap. 8), en el que todos los participantes deben cumplir las condiciones de:

- No estar obligados.
- Tener la posibilidad de defender su postura.
- Aceptar las condiciones bajo las que se ha establecido el acuerdo.
- Negociar, que no consiste simplemente en «una coincidencia entre las opciones individuales» (1980, pág. 165).
- No prestar excesiva atención a los propios intereses.
- Adoptar un acuerdo que afecte a todos por igual.
- Seleccionar una política para la evaluación que sirva a los intereses del grupo al que se aplica.
- Tener una información completa e igual sobre los hechos relevantes.
- Evitar que unas evaluaciones incompetentes y arbitrarias supongan algún riesgo para los participantes.

House defiende esta postura reformista. Como respuesta a los críticos que le tachan de tendencioso a favor de los desfavorecidos, dice: «Me parece que asegurarse de que se representen los intereses de los desfavorecidos y se consideren en serio no significa ser tendencioso; al contrario, sin duda es una postura más igualitaria que la mayoría de las prácticas actuales» (1991a, págs. 241-242).

La justicia constructivista. El cambio constructivista de la sociología plantea unas cuestiones especiales referentes a la justicia. Johnson y Altheide (Altheide y Johnson, 1997; Johnson, 1995) han realizado algunos trabajos sobre la justicia desde una perspectiva reflexiva. Rechazan las teorías actuales de justicia, y Johnson sostiene que, según una teoría reflexiva, la justicia:

- Es una lucha: «Nuestra lucha por vivir con virtud en nuestras vidas sociales y comunales» (Johnson, 1995, pág. 199).
- Es fundamentalmente emocional. Nuestros juicios nacen de nuestra experiencia.

- Es evolutiva, cambia con el tiempo y cambian los criterios a medida que cambiamos de tiempo y lugar.
- Tiene en cuenta los sexos. Entre nuestras batallas está «una preocupación permanente por cómo vivir en paz y de forma productiva con las personas del otro sexo» (1995, pág. 201).
- Es personal: «La búsqueda de la justicia no está en nuestras palabras ni en nuestras teorías, sino en nuestras acciones concretas para ayudar a las otras personas» (ibíd.).
- Es una cuestión de interacción social: «Lo habitual es que nos unamos a los demás [...] La justicia es comunal y procesal» (ibíd.).
- Es nuestro esfuerzo por «entender qué ocurre en el mundo» (ibíd.).
- Es egoísta: «El egoísmo es el fundamento del altruismo» (1995, pág. 203).

Este planteamiento nos lleva a una preocupación empírica por «cómo se constituye, se comunica y se experimenta la justicia en la vida diaria» (Altheide y Johnson, 1997, pág. 184). Las críticas de las posiciones constructivistas e interpretativistas se centran sobre todo en su relativismo y, por consiguiente, en su potencial de autocomplacencia política. Greene concluye:

Las cuestiones sobre los objetivos políticos y morales del interpretativismo son las más profundamente inquietantes. El relativismo esencial del interpretativismo defiende la no existencia de ningún papel particular en el mundo. No existe el compromiso [...] de reforzar el *statu quo* ni de ponerlo en entredicho, de aprobar las prácticas racistas o sexistas con los propios silencios o de intentar reparar las desigualdades de forma activa, de permanecer neutral en cuestiones de creencias políticas o de defender abiertamente una ideología determinada (1992, págs. 43-44).

DISEÑAR LA EVALUACIÓN CUALITATIVA PARA LOS RESULTADOS

Los estudios cualitativos no están diseñados para proporcionar respuestas definitivas a preguntas causales [...] [pero] puede seguir siendo una pretensión con las debidas reservas.

LOFLAND y LOFLAND (1995, págs. 136, 138).

Miles y Huberman son menos reservados aún: «La idea convencional es que los estudios cualitativos sólo son buenos para incursiones exploratorias, para desarrollar hipótesis, y que las explicaciones sólidas, incluidas las atribuciones causales, sólo se pueden derivar de estudios cuantitativos». Dicen de esta idea que es «falsa» (Miles y Huberman, 1994, pág. 147), e insisten en que la investigación de evaluación cualitativa:

- Identifica los mecanismos causales.
- Se ocupa de las complejas redes locales.
- Pone en orden la dimensión temporal de los acontecimientos.
- Está preparada para avanzar y retroceder entre los diferentes niveles de variables y procesos.
- La inducción analítica es una forma de comprobar las explicaciones de los casos concretos y de profundizar en ellas.

La evaluación cualitativa puede abordar los problemas de resultados y las cuestiones de atribución causal de tres formas. En primer lugar, hay soluciones de diseño análogas a los diseños que implican un grado de control. Se inspiran sobre todo en la obra de Donald Campbell. Otros autores, como Cronbach y Yin, que abogan por el uso del estudio múltiple de casos, han adoptado soluciones de diseño de esta misma amplia familia. En segundo lugar, el cambio de énfasis desde la validez interna a las cuestiones de la validez externa y de la generalización se ha traducido en una mayor sensibilidad hacia los microprocesos de la práctica y los programas. En tercer lugar, el impacto del realismo científico ha llevado a una reformulación radical de las ideas de causalidad. Hablaremos de todo ello.

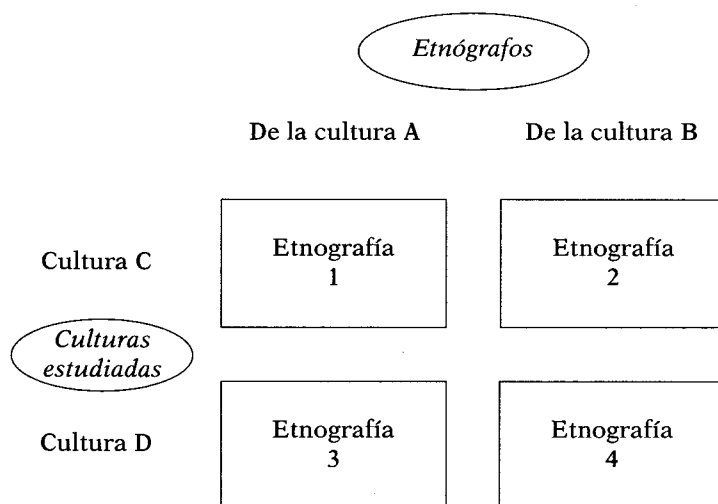
Campbell y el estudio de casos

En sus primeros tiempos, Campbell decía que los diseños de estudio de casos «de efecto inmediato» no se pueden interpretar en lo que a la atribución causal se refiere. Sin embargo, a través de los intercambios con Becker y Erikson, vino a adoptar la teoría de que la analogía de los grados de libertad es una fuente importante de disciplina para interpretar los estudios de casos intensivos y transculturales. La lógica conlleva comprobar las teorías de acuerdo con las predicciones o las expectativas que estimula, a través de un proceso que

el autor describe como correspondencia de patrones, basado en la premisa de que «el diseño experimental se puede separar de la cuantificación» (Campbell, 1978, pág. 197).

Campbell (1964) sugiere un paradigma que implica un compromiso incondicional tanto con la etnografía como con los estudios comparativos (véase la figura 7.1). En este modelo, dos antropólogos de diferentes culturas estudiarían, cada uno de ellos, una tercera y una cuarta cultura. Los efectos globales se «leen» de la siguiente forma:

- Los atributos comunes a las etnografías 1 y 3, pero que no se comparten con la 2 y la 4, se pueden atribuir al etnógrafo A.
- Los atributos comunes a las etnografías 2 y 4, pero que no son evidentes en ningún otro lugar, se pueden atribuir al etnógrafo B.
- Los atributos comunes a las etnografías 1 y 2, pero que no están presentes en la 3 y la 4, se pueden atribuir a la cultura C como conocidos «objetivamente».
- Los atributos comunes a las etnografías 3 y 4, pero que no se encuentran en la 1 y la 2, se pueden atribuir a la cultura D como conocidos «objetivamente».



Fuente: Campbell (1964).

FIGURA 7.1. Diseño de multi-etnografía para sacar el contenido de la cultura estudiada y al que el etnógrafo ha contribuido.

Los atributos comunes a *las cuatro etnografías* son inherentemente ambiguos y no se pueden interpretar ni como una tendenciosidad de los etnógrafos, ni como características compartidas por ambas culturas. Las características interpretativas *exclusivas* de cualquier etnografía son indeterminadas, debido a la especificidad del etnógrafo, la cultura o las interacciones entre ambos.

La aplicación completa de este modelo requeriría un uso intensivo de recursos, lo que no es viable. No obstante, sería un gran avance operar según la mitad superior de la figura. Otra variante consistiría en que cada etnógrafo estudiara la cultura del otro, con unas bases de referencia implícitas facilitadas por las descripciones que los etnógrafos hicieran de sus propias culturas. En este modelo sería deseable disponer de etnógrafos de culturas tan diferentes entre sí como fuera posible.

Los microprocesos

Veámos en el capítulo 2 que a William Reid le atraía el potencial de la investigación de «proceso de cambio». El autor no rechaza el papel de los experimentos controlados, pero su conclusión es que «las limitaciones prácticas y éticas de los experimentos exigen una dependencia del estudio naturalista de estas relaciones» (Reid, 1990, pág. 130). Esto implica centrarse en los procesos de cambio durante el período de contacto entre el ayudante profesional y el sistema del cliente. Este planteamiento, en vez de confiar en mediciones acumuladas y promediadas de los resultados, vuelve a la tradición del análisis de contenidos de la investigación social, prestando una mayor atención a los microrresultados.

Reid aplica sus ideas a la asistencia social, aunque la lógica podría aplicarse a otras formas de servicios profesionales orientados al cambio. En la base de su teoría se observa una idea sistémica de la intervención, según la cual el profesional y el usuario del servicio se encuentran en una interacción circular y mutuamente influyente. En este modelo, «las distinciones convencionales entre intervención y resultados pierden su simplicidad» (1990, pág. 135). «Entonces se hace posible representar la investigación de proceso de cambio como el estudio de unas cadenas en las que se entrelazan las intervenciones y los resultados» (1990, pág. 136). Reid pretende defender los experimentos, pero apunta a una postura más naturalista cuando dice que «las medias de las variables del proceso que carecen de sus contextos propor-

cionan, en el mejor de los casos, unas mediciones deficientes» (1990, pág. 137).

Este enfoque resulta especialmente prometedor en los estudios de casos cuya unidad de análisis sea un individuo y donde el evaluando incluya la prestación de un servicio de uno a uno. Evidentemente, la evaluación de microprocesos no se limita a la evaluación orientada a los resultados. En el capítulo 3 señalábamos la crítica que Silverman hace de algunas formas de investigación de evaluación que poseen cierto prestigio. El autor ha aplicado el análisis de la conversación a la evaluación de la organización secuencial de las charlas en el asesoramiento sobre el sida. Sostiene que el análisis de la conversación es apropiado para comprender este tipo de datos complejos, pero también que puede establecer de forma demostrable un diálogo provechoso con los profesionales en toda una amplia variedad de enclaves (Silverman, 1997a, pág. 37).

Causa y modelos causales

La tercera formulación cualitativa de los problemas metodológicos que plantea la evaluación de resultados es un tanto diferente. Parte del estímulo que proporciona otro tipo de razonamiento referente a la naturaleza de la causa, y los correspondientes modelos de hipótesis causal que surgen de tal razonamiento. En el capítulo 3 esbozábamos el trasfondo de estos cambios al referirnos a la llegada de las explicaciones realistas científicas y a su nueva aplicación a la evaluación. La idea central es la de unos mecanismos causales subyacentes que no se pueden comprender con trabajos y mediciones superficiales. Por consiguiente, «los propios acontecimientos no son el objetivo último del análisis científico [...] La realidad no sólo consiste en lo que se puede ver, sino también en las entidades causales subyacentes que no siempre se pueden discernir» (House, 1991b, pág. 4). La realidad oculta produce unos acontecimientos reales, de los que tenemos experiencias empíricas e impresiones sensitivas.

Tal idea se describe a menudo como un concepto *generativo* de la causalidad:

Cuando damos una explicación generativa de un resultado, no presentamos unas variables o unos correlatos que se asocien entre sí; al contrario, intentamos explicar cómo se produce la propia asociación.

Los mecanismos generativos, por tanto, realmente *constituyen* el resultado (Pawson y Tilley, 1997b, pág. 408).

El concepto convencional de la causalidad como regularidades y asociaciones se rechaza en favor de unas entidades causales que tienen unas «tendencias que interactúan con otras tendencias, de tal forma que se puede producir o no un acontecimiento observable» (House, 1991b, pág. 5). House cita a Manicas y a Secord cuando dice que «para la visión estándar de la ciencia, el mundo es una concatenación determinada de acontecimientos contingentes; para la realista, es una concatenación contingente de estructuras reales. Y esta diferencia es monumental» (House, 1991b, pág. 5).

En consecuencia, el realista, en vez de limitarse a documentar la secuencia y la asociación de los acontecimientos, pretende *explicar* estos acontecimientos. Aunque esta idea de causa no exige necesariamente una metodología cualitativa, se presta sin duda a este tipo de métodos. Las explicaciones causales serán locales y «orientadas al ahora» (Lofland y Lofland, 1995, pág. 141). Miles y Huberman desarrollan unos métodos analíticos que abordan la atribución causal tanto en las explicaciones de casos múltiples como en los singulares. Por ejemplo, abogan por el uso de la investigación de campo para trazar el mapa de las «redes causales locales» que los informantes llevan en la cabeza y para establecer conexiones con el propio mapa causal emergente que el evaluador traza del enclave en cuestión. Estos mapas parten de unos «fragmentos causales» que conducen a la construcción de unas cadenas lógicas de pruebas. Tales redes causales «no son probabilísticas, sino específicas y determinadas, basadas en la comprensión de los acontecimientos a lo largo del tiempo en el contexto local correcto —y *vinculadas* a una buena conceptualización de cada variable—» (Miles y Huberman, 1994, pág. 159).

Cronbach adelantó gran parte de este razonamiento en sus tesis sobre los modelos causales. Al rechazar la idea de causalidad como unos acontecimientos que se pueden predecir con un alto grado de probabilidad, Cronbach desarrolló unas tesis gemelas. En primer lugar, decía que las causas son contingentes en lo que a las interacciones locales de los grupos de acontecimientos se refiere. En segundo lugar, aceptaba que normalmente existen unos sucesos o unas condiciones ausentes que afectan al resultado de un determinado programa, pero de los que sabemos muy poco. Fue el primer teórico que dio una explicación verosímil de los factores contextuales de la evaluación. De ahí concluye que «después de que quien realiza el experi-

mento, con su coacción artificial, abandona la escena, es seguro que el programa en funcionamiento se adaptará a las condiciones locales» (Cronbach y otros, 1980, pág. 217). Además, «la evaluación de un programa depende tanto de su contexto que la duplicación no es más que una figura retórica» (1980, pág. 222).

En su conjunto, la intervención experimental no es ni necesaria ni suficiente para que se produzca el efecto previsto. Un experimento no puede proporcionar una prueba decisiva de la efectividad de un programa. La formulación tradicional «no es el mundo de los programas sociales y, en general, no es el mundo social en absoluto» (House, 1993, pág. 135).

LOS ESTUDIOS DE CASOS EN LA EVALUACIÓN

La justificación de los estudios de casos en las evaluaciones cualitativas se sitúa en la confluencia de su idea receptiva de valor político y de sus supuestos interpretativistas subyacentes.

GREENE (1994, pág. 538).

Las fuertes críticas a los diseños experimentales tradicionales y los desarrollos de la metodología cualitativa en los que se puede incorporar una interpretación parcial de los procesos causales se han unido para situar en primer plano a los estudios de casos como un diseño de evaluación que se centra en los resultados de la evaluación junto con una apreciación de la naturaleza contingente y constructivista de las políticas, los programas y las prácticas.

Los estudios de casos en la evaluación surgieron en los años setenta en el ámbito de la educación. «Estudio de casos era una imagen que atraía a quienes buscaban una forma de integrar la exigencia de unos datos comprensivos que se planteaba en diversas revisiones críticas de la tradición de la evaluación» (Simon, 1987, págs. 61-62). En el contexto de la evaluación, el estudio de casos no fue un simple préstamo de las tradiciones de la ciencia social existentes que se remontan a la Escuela de Chicago. Para bien o para mal, a los evaluadores les preocupaban los temas relacionados con el público, las limitaciones de tiempo y las relaciones de poder, que producían un modelo cooperativo de investigación. Muchos de los estudios de casos realizados en ese período «carecían al menos de tres características clave de la investigación etnográfica: la inmersión en el campo, la observa-

ción directa como técnica dominante y una teoría interpretativa del caso por parte del etnógrafo» (Simon, 1987, pág. 79). Simon piensa que los estudios de casos de los años setenta y ochenta se acercaban más a la historia oral que a la etnografía. Las diferencias siguen existiendo en estos momentos. Si, por ejemplo, se comparan los textos sobre estudios de casos del teórico de la evaluación Stake (1995) y del metodólogo Yin (1994), se observa una gran variación en el énfasis, reflejo de esas diferencias.

Hakim señala que «el estudio de casos es la investigación social equivalente al reflector o al microscopio; su valor depende fundamentalmente de lo bien que se enfoque el estudio» (1988, pág. 61). Los estudios de casos son flexibles y de propósito múltiple. Pueden ser descriptivos, cuando exploran entidades poco conocidas y facilitan imágenes de ellas. También pueden ser selectivos, cuando pretenden dar una explicación más rica y detallada de los procesos en acción. Se pueden diseñar también para conseguir una forma de aislamiento experimental de los factores sociales seleccionados dentro de un contexto de la vida real. Pero cualquiera que sea la finalidad general que guíe la adopción de la evaluación basada en el estudio de casos, las decisiones sobre el diseño del estudio deben tener en cuenta seis cuestiones. Primero, ¿qué se considera un «caso»? Segundo, ¿qué criterios se deben aplicar en la decisión de si un programa o un proyecto se ha de evaluar o no? Tercero, ¿la evaluación abordará un caso único o múltiples casos? Cuarto, ¿cuáles serán las unidades de análisis? Quinto, ¿qué lógica de generalización se va a adoptar? Y por último, ¿cuál es la relación que se pretende establecer entre los valores del profesional y los del evaluador? Nos ocuparemos brevemente de cada una de estas preguntas.

En primer lugar, ¿qué se considera un caso? Dice Stake:

El caso es un algo especial que se ha de estudiar: un alumno, una clase, una comisión, un programa, quizá; pero no un problema, una relación ni un tema. El caso que se ha de estudiar probablemente tiene problemas y relaciones, y es previsible que el informe del caso tenga un tema, pero el caso es una entidad. El caso, en cierta forma, tiene una vida exclusiva (1995, pág. 133).

En el capítulo anterior señalábamos las críticas exageradamente generalizadas y tan inoportunas que Atkinson y Delamont hacen de los evaluadores de estudios de casos, pero estos autores no siempre yerran el tiro. Por ejemplo, se quejan de una tendencia a la falta de rigor cuando se habla de los «casos»:

Determinar qué constituye un «caso» es [...] mucho más problemático de lo que parecen aceptar los investigadores de «estudio de casos» [...] No tiene [...] sentido alguno que los autores convencidos de los estudios de casos escriban como si el mundo estuviera poblado de «casos» cuyo estatus y cuya existencia fueran independientes de los intereses metodológicos y teóricos (Atkinson y Delamont, 1993, pág. 207).

La definición tan cauta de Stake demuestra en cierto grado esta vaguedad. Pero lo anterior sugiere que los estudios de casos se pueden realizar con individuos, comunidades, grupos sociales, organizaciones e instituciones, y determinados acontecimientos, roles y relaciones o interacciones (Hakim, 1988).

Segundo, al decidir si hay que evaluar o no un programa o un proyecto determinados, Cronbach establece, como principal criterio, la reducción de la incertidumbre y la producción de información: «El *poder de acción* es la base. El poder de acción se refiere a la probabilidad de que la información, si se cree, cambie el curso de los acontecimientos» (Cronbach y otros, 1980, pág. 265). No hay que interpretar estas palabras como una idea instrumentalista del uso de la evaluación porque «las preguntas que mayor poder dan a la acción son las que contribuyen a la reflexión» (1980, pág. 266). El autor propone un simple algoritmo:

1. Las preguntas con un elevado poder de acción y una elevada incertidumbre merecen que se les preste mucha atención.
2. Un menor poder de acción, sin que deje de existir, más una menor incertidumbre exigen cierta inversión.
3. En las situaciones de mucho poder de acción y poca incertidumbre, se deben emplear métodos de bajo coste.
4. En las situaciones de poco poder de acción y poca incertidumbre no se debe realizar evaluación alguna.

Sin embargo, las cuestiones del «poder de acción» y de la «incertidumbre» seguirán abiertas a nuevas interpretaciones.

En tercer lugar, los estudios de casos pueden ocuparse de casos únicos o de casos múltiples. Hay varias razones para centrarse en un caso único. Yin distingue entre casos críticos (que permiten la comprobación de una teoría bien formulada), casos extremos o raros, y casos reveladores (un fenómeno de momento inaccesible para la investigación científica). Por ejemplo, las organizaciones benéficas pueden poner en marcha proyectos que reflejen una nueva forma de in-

tervención que ellos desean desarrollar; las autoridades educativas locales pueden introducir pequeñas innovaciones curriculares aunque sean relativamente costosas; y los departamentos gubernamentales implantan de manera rutinaria formas innovadoras o importadas de tratar a los infractores. En estos casos son posibles los estudios de casos críticos o raros. Cambiar las actitudes públicas hacia los beneficios o los riesgos de un proyecto hasta la fecha intachable puede abrir la posibilidad de un estudio de casos revelador en un área de la práctica previamente prohibida.

El ámbito de las innovaciones suele ser muy extenso y, por consiguiente, cabe la posibilidad de unos estudios de casos múltiples, y no sólo únicos. Las innovaciones escolares suelen estar en esta categoría. Por definición, no es probable que sean casos-críticos, raros ni reveladores.

Los estudios de casos múltiples se pueden conseguir mediante «proyectos diseñados específicamente como una serie de estudios etnográficos en diferentes enclaves, seleccionados según los criterios desarrollados a partir de la teoría existente para proporcionar las dimensiones más significativas para la comparación» (Finch, 1986, pág. 185). La autora señala la ventaja de contar con un equipo central que marque una dirección para los temas y las categorías analíticos en la fase de diseño, pero que no imponga los métodos de investigación. Otra posibilidad es que los estudios de caso se realicen de forma secuencial, mediante la planificación de unos estudios comparativos acumulativos. Aunque no hay que esperar que esto produzca un fácil desarrollo lineal de los conocimientos, sí que abre la oportunidad de una evaluación interdisciplinar, y sería un avance significativo en la mayor parte de las evaluaciones basadas en el estudio de casos actuales.

Queda claro en este ejemplo que a los autores les interesaban los casos en más de un nivel. Parte de su argumentación se ocupa de los análisis holísticos de cada centro escolar, pero parece que les interesan también las diferencias de clase social, de política de admisión de alumnos y de ubicación, urbana o rural. Yin emplea el término «incrustado» para referirse a los estudios de casos en los que hay más de una unidad de análisis, y éste es el cuarto aspecto de la decisión sobre el diseño que los evaluadores deben resolver.

La quinta cuestión referente al diseño del estudio es la lógica de la generalización a partir de los estudios de casos. El problema de la generalización será distinto según si los estudios de casos tienen un objetivo principalmente descriptivo o explicativo. En cualquier caso, Yin

advierte de que es inadecuado recurrir a la lógica de la encuesta cuando se planifiquen estudios de casos, y subraya la distinción entre generalización analítica y generalización estadística. «Un error fatal en la realización de estudios de casos es pensar en la generalización estadística como método para generalizar los resultados del caso» (Yin, 1994, pág. 31). Los problemas de la validez externa en la evaluación basada en el estudio de casos y, más en general, en la evaluación cualitativa, se confunden por establecer analogías con muestras y poblaciones. Por consiguiente, cuando se seleccionen casos múltiples, la lógica no es la de la muestra, sino la de la réplica. «Cada caso se debe seleccionar de manera que: o a) prediga unos resultados similares (una *réplica literal*), o b) produzca unos resultados opuestos pero por razones predecibles (una *replica teórica*)» (Yin, 1994, pág. 46).

EJEMPLO 7.3. EL ESTUDIO DE CASOS EN ENCLAVES MÚLTIPLES

Burgess y otros (1994) exponen un estudio de casos de enclaves múltiples sobre la introducción de registros de rendimientos en los centros de enseñanza secundaria de un condado de la región central de Inglaterra. Un investigador (dos años) y dos profesores de apoyo (un año cada uno) pasaron un trimestre en uno de los centros, trabajando con los profesores y los alumnos y, siempre que era posible, participando en las actividades relacionadas con el proceso de registro de rendimientos.

Los cuatro centros eran «ilustrativos de la variedad de centros que existían en el distrito» (Burgess y otros, 1994, pág. 131) y permitían hacer comparaciones entre centros urbanos y rurales, centros selectivos y no selectivos, centros con alumnos procedentes predominantemente de familias de clase trabajadora y otros con alumnos hijos de familias de clase media, y centros con todos sus alumnos blancos y otros de alumnado multiétnico. Los centros también diferían en el grado en que se había introducido el programa. Los autores dicen: «En general, la virtud del estudio estaba en su carácter comparativo. El diseño de la investigación facilitó el examen de una amplia variedad de temas educativos sustantivos en diferentes contextos» (1994, pág. 136).

Se elaboraron informes individuales sobre cada centro, además de un informe temático. Se abordaron tanto las cuestiones temáticas como la importancia del tiempo, como los temas específicos de cada centro en particular. Los autores destacan la interconexión de los recursos, la selección de los sitios y el análisis, y también hablan implí-

citamente de algunos peligros de otorgar demasiado peso al contexto local. Por ejemplo, dicen que el hecho de plantear a los centros temas similares al principio de la evaluación abrió la posibilidad de un análisis entre los diferentes casos.

La argumentación de Yin es sólida. Pero debemos señalar que los modelos de generalización estadística se han utilizado de forma eficaz en algunas ocasiones con estudios de casos múltiples. Sinclair (1971) habla de un estudio sobre centros para jóvenes delincuentes en el que se utilizó una variedad impresionante de datos para sacar conclusiones referentes al efecto que los entornos del centro producían en la delincuencia. Decía con otros autores que:

La comparación de un gran número de instituciones dentro de un único estudio permite superar un defecto grave de los estudios de casos descriptivos. Se trata de la dificultad de vincular las características exactas de la atención prestada con sus efectos inmediatos y a largo plazo (Tizard y otros, 1975, pág. 3).

Los autores exponen una serie de estudios sobre centros residenciales con diferente ideología, organización, personal y conducta de los internos. El riesgo de este tipo de evaluaciones cualitativas es recurrir a minietnografías superficiales. Se han realizado muchos estudios de casos que se basan en un trabajo de campo que no duró más de dos o tres días. Pero, si los recursos son los adecuados, los estudios de casos abren muchas oportunidades para una evaluación cualitativa teóricamente informada.

En sexto lugar, señalábamos la observación de Simon de que la aparición de los estudios de casos en el mundo de la educación incluía una preocupación por las relaciones de poder, y fomentaba el surgimiento de los modelos colaborativos de estudios de casos. Los modelos de evaluación democrática de Gran Bretaña y Estados Unidos se desarrollaron en un contexto político que respaldaba los enfoques de una educación que buscara el descubrimiento. Esto favoreció que pronto se reconociera que los estudios de casos pueden tener un potencial emancipador «basado en el respeto por la autonomía del individuo y el principio de la automejora estimulada» (Simon, 1987, pág. 83). En otras palabras, se consideraba que el estudio de casos poseía en sí mismo un potencial educativo. Los estudios de casos educativos realizados en Gran Bretaña en los años setenta y ochenta incluí-

an la creencia en que «las personas son dueñas de los actos de su vida» y llevaron a los evaluadores «a querer mostrar la divergencia, más que resolverla» (1987, pág. 86). Los estudios de casos que son ampliamente emancipadores también se han desarrollado en otros campos.

EJEMPLO 7.4. EL ESTUDIO DE CASOS PARA LOS PROFESIONALES Y LOS INVESTIGADORES

Bogdan y Taylor (1994) exponen un estudio de casos de enclaves múltiples de cuarenta programas, cuyo fin era el fomento de la integración en la comunidad de personas con dificultades de aprendizaje y discapacidades múltiples. Los autores pretendían desarrollar un modelo de evaluación que resolviera en parte las orientaciones opuestas de los profesionales y de los investigadores, combinando para ello adecuadamente el rigor escéptico con una investigación que «ayudara a los profesionales conscientes, a las personas que encabezan las reformas que tienden a la integración» (1994, pág. 295).

Estudiaron cada uno de los cuarenta proyectos mediante un programa renovable de ocho breves estudios de casos al año, durante cinco años. Los enclaves se seleccionaron mediante métodos de muestreo intencional en los que se pedía a los informantes que propusieran un buen programa y se realizaban entrevistas por teléfono. Los autores pretendían identificar los programas innovadores y ejemplares, más que elaborar una muestra de probabilidad: «Conscientemente intentábamos encontrar lugares que nos pudieran enseñar algo sobre cómo se pueden integrar en la comunidad las personas con discapacidades graves» (1994, pág. 298). Coherentes con su orientación teórica, empezaron con una definición intencionadamente vaga de la integración:

Tratamos el concepto de integración como problemático, algo que se ha de investigar más que asumir. Queremos descubrir de qué modo los organismos [...] definen y consiguen la integración (Bogdan y Taylor, 1994, pág. 297).

El resultado de su planteamiento fue un acceso más fácil y, paradójicamente, las personas, al hablar de los dilemas que se les planteaban en la práctica, eran más sinceras de lo que hubieran sido en otras circunstancias. Se publicaron informes breves sobre cada proyecto, y cuando esos informes se centraban en los aspectos negativos de organismos nada ejemplares, a éstos se les permitía decidir si se mencionaba o no su nombre.

A los autores les interesaba la generalización —«unos patrones que trasciendan de los casos individuales» (1994, pág. 300)—. También les interesaba teorizar y «desarrollar unos conceptos sintetizadores y una teoría asentada que trascendieran las ideas de sentido común de las personas que estudiamos» (ibíd). No decían cómo pensaban que funcionaba la generalización, ni hacían reflexión alguna sobre la relación establecida entre los conceptos profanos y los científicos.

LA ACCIÓN Y LA PARTICIPACIÓN

La investigación en la acción «como una forma de generar conocimientos acerca de todo un sistema social y, al mismo tiempo, tratar de cambiarlo» (Elden y Chisholm, 1993, pág. 121) se ha asociado con la obra de Lewin y, en Gran Bretaña, con la del Instituto Tavistock, centrada en los problemas internos de las organizaciones y de la vida laboral. A finales de los años sesenta, la investigación en la acción se asociaba a algo muy distinto, a través del Programa Americano sobre la Pobreza, la introducción de los programas de discriminación positiva en la educación y los Proyectos de Desarrollo Comunitario a través del Ministerio del Interior en Gran Bretaña. No se pensaba que la evaluación versara principalmente sobre el desarrollo disciplinar, sino que se pretendía que respondiera directamente a las necesidades y los planes que estableciera el programa de acción. En muchos casos, aunque no en todos, representaba una idea instrumental y activista de la investigación, como una empresa intrínsecamente política.

La denominación se ha conservado en el ámbito de la salud, el del bienestar y el de la educación (Hart y Bond, 1995). En la educación abarca toda una diversidad de actividades, desde la autoevaluación hasta la investigación colaborativa en grupo (Ebbutt, 1985; Kelly, 1985).

Carr y Kemmis ofrecen pocas orientaciones explícitas sobre el diseño de la investigación en la acción. La investigación y la evaluación de las diferentes formas de participación del ciudadano han materializado algunas áreas de buena práctica, basadas en la creencia de que «después de todo, estudiar la participación de forma excluyente y que no permita esa misma participación supone una incoherencia mayúscula» (Beresford, 1992, pág. 18). Gran parte de esta buena práctica está relacionada con los términos del acuerdo de evaluación justa que señalábamos antes en este capítulo, cuando hablábamos de

la justicia como imparcialidad. Este consejo se puede resumir como hacemos a continuación.

En primer lugar, aclarar y acordar mutuamente el *propósito de la participación*. Por ejemplo, los gestores de organismos liberales pueden entenderla como un medio para mejorar la receptividad del servicio, mientras que los usuarios del servicio pueden considerarla como un medio de potenciación del usuario, «para ampliar la capacidad de los usuarios para participar en las decisiones relacionadas con el diseño, la gestión y la revisión de los servicios» (Barnes y Wistow, 1992, pág. 4).

En segundo lugar, ser claros en lo que se refiere a los diferentes *tipos de «usuarios»* que puedan estar interesados. Fisher, por ejemplo, indica que gran parte de la evaluación participativa peca de ingenuidad cuando menosprecia las diferencias que existe entre los clientes voluntarios y los «cautivos». «Un problema del nuevo lenguaje voluntarista de la protección al consumidor es que se reconoce muy poco a los usuarios que están “cautivos”» (Fisher, 1992, pág. 55). Esto lleva a los trabajadores sociales a presumir erróneamente que existe un acuerdo con los clientes cuando no se ha alcanzado tal acuerdo: «Forma parte del equipamiento profesional suponer que el trabajo se basa en la negociación y el acuerdo» (1992, pág. 48). Tales acuerdos tienden a ser coercitivos (es voluntario siempre que uno esté de acuerdo); condicionales (porque su finalidad es obtener algún resultado de la asistencia social); y limitados (por los problemas de recursos que limitan la auténtica elección).

EJEMPLO 7.5. LA CIENCIA EDUCATIVA CRÍTICA

Carr y Kemmis (1986) defienden la investigación en la acción como «ciencia educativa crítica» y, en consecuencia, la sitúan decididamente dentro de la evaluación crítica. Delimitan cinco características. Primera, se basa en una idea dialéctica de la racionalidad. Como tal, implica la transformación de las prácticas a lo largo del tiempo y la búsqueda de correspondencias y no correspondencias entre las prácticas y las interpretaciones; las prácticas y las situaciones; y las interpretaciones y las situaciones. La investigación en la acción es deliberadamente activista. Los autores rechazan el «determinismo y el fisicalismo» del objetivismo y el «relativismo y la teoría racionalista de la acción» del subjetivismo (Carr y Kemmis, 1986, pág. 184).

Segunda, emplea las categorías interpretativas del profesor. No obstante, rechazan con firmeza la idea de que la «teoría tácita» representa una variedad del «tener una teoría»:

Afirmar que la teoría es algo que uno puede defender «implícitamente» o inconscientemente supone un uso erróneo de la idea de «teoría». En efecto, para que la idea de «teoría» tenga alguna fuerza es necesario que sea algo que se adopta conscientemente [...] y que sea el producto de la reflexión, no del hábito, la costumbre o la coerción (1986, pág. 189).

El objetivo de los autores es la «acción informada y comprometida» de la praxis. «Por lo tanto, la investigación en la acción no puede ser más que la investigación de la propia práctica» (1986, pág. 191).

Tercera, la investigación en la acción sitúa a la falsa conciencia. Con ello, también «trata al actor como el poseedor de la ideología además de la "víctima"» (1986, pág. 193). Cuarta, la investigación en la acción irá de la organización de la ilustración a la organización de la acción, a través de la exposición de los aspectos del orden social que frustran el cambio. En este sentido, los autores insisten en las dimensiones colectiva y colaborativa de la autorreflexión y en los peligros de la autorreflexión de un sujeto solo.

Por último, la investigación en la acción une la teoría y la práctica. Los autores rechazan, por su «radicalismo irresponsable», la tendencia de algunos teóricos críticos a sospechar demasiado de la utilidad.

Fisher subraya también la importancia de las diferencias entre los usuarios del servicio a corto y largo plazo, y los usuarios actuales y potenciales, y dice que hay que esforzarse en reconocer los temas organizativos y de recursos que puedan enfrentar a los usuarios entre sí.

En tercer lugar, establecer unos mecanismos auténticos que aseguren un trabajo colaborativo continuado. Los acuerdos verbales suelen pecar de poco concretos y tienden a deteriorarse de forma crónica.

Cuarto, conseguir un acuerdo sobre si el programa o proyecto es evaluable. Nos ocupábamos de este tema al principio de este capítulo, cuando hablábamos de las valoraciones de la evaluabilidad. Los problemas concretos que hay que destacar son los riesgos recurrentes de intromisión e imposición: «Tanto si la intromisión en nuestra vida se hace alegando que se "ayuda" como si se hace aduciendo que se "investiga", parece que la experiencia es muy similar cuando uno ocupa la posición del destinatario» (Davis, 1992, pág. 37).

Por último, ponerse de acuerdo sobre los perjuicios que deben evitarse y los beneficios que han de conseguir quienes colaboren. Barnes y Wistow (1994), al reflexionar sobre un proyecto extensivo para implicar a los cuidadores en los procesos de planificación y de revisión del

Departamento de Servicios Sociales de una autoridad local inglesa, advierten de que existe el peligro de meditar poco si el hecho de satisfacer las necesidades del sistema de planificación del servicio y el de satisfacer las necesidades de los usuarios y los cuidadores que se ven involucrados en ese sistema son incompatibles:

Quienes prestan el servicio deben ser sensibles a las exigencias que plantean a las personas, que ya tienen una vida complicada y difícil. Han de poder garantizar que las personas saquen algún beneficio directo de la participación, y asegurar también que [...] la participación se justifique por los resultados que con ella se consigan (1994, pág. 91).

Hemos analizado la lógica y los valores de la evaluación cualitativa, y de qué forma configuran las decisiones referentes al diseño de la evaluación. Hemos considerado las diferentes formas en que se puede valorar la evaluabilidad de un proyecto o un programa. Los temas del consentimiento y de la intimidación sirvieron de base para hablar de los problemas éticos que se plantean en el diseño de la evaluación. Perfilamos de qué forma los evaluadores pueden tener en cuenta ideas nuevas sobre la importancia de los fines de justicia social.

La investigación de evaluación cualitativa puede abordar los temas de los resultados; la argumentación de tal idea ha sido un punto central de este capítulo. Los estudios de casos pueden ocuparse de estos temas, entre otros, y hemos revisado las distintas cuestiones que el estudio de casos debe tener en cuenta. El capítulo concluía con unas reflexiones sobre las consideraciones referentes al diseño que plantean los avances de la evaluación participativa y orientada a la acción.

Iniciamos el capítulo siguiente con una introducción a los criterios que rigen en las decisiones prácticas de diseño en lo que a los métodos de recogida de datos se refiere. Las entrevistas individuales y de grupo, la observación, las historias vitales, los documentos oficiales y vivos y las simulaciones forman parte del repertorio tanto de la evaluación como de la investigación cualitativas. El propósito evaluativo es, una vez más, lo que da al trabajo de campo de la evaluación su sello particular.

CAPÍTULO 8

LA EVALUACIÓN EN EL CAMPO

Si estábamos en lo cierto cuando a lo largo del libro hemos estado diciendo que la indagación es evaluativa cuando está marcada por el propósito evaluador, entonces las formas que adopte el trabajo de campo son asimismo contingentes según sea el propósito. Los métodos no serán ni enteramente los mismos que los de la investigación cualitativa en general, ni completamente distintos. Por esta razón, soy reacio a rechazar de plano *cualquier* método de indagación. Pero tampoco recomendaría que se adopten de forma rutinaria los dispositivos metodológicos, aunque se trate de los que en apariencia resultan más relevantes. El resultado del automatismo metodológico será unas evaluaciones forzadas y torpes, carentes de rigor y de relevancia.

Cualquier buen texto sobre metodología de la evaluación cualitativa se puede traducir a una serie de dialectos que constituyen los propósitos de la evaluación. Unos están escritos con la mirada puesta en esa traducción (Bogdan y Taylor, 1994; Crabtree y Miller, 1992a; Holstein y Gubrium, 1995). Como su carácter prescriptivo es menos evidente, éstos son al menos tan útiles para el evaluador como los escritos explícitamente sobre la evaluación cualitativa (Patton, 1990), o sobre la evaluación en general (Wholey, 1994).

Mi objetivo en este capítulo es ofrecer una introducción ilustrativa a una perspectiva evaluativa sobre el trabajo de campo, más que un inventario resumido de métodos de trabajo de campo que en otras partes de la literatura se tratan con mayor detalle. En la parte principal del capítulo analizaremos de qué forma los propósitos evaluativos configuran las decisiones de emplear entrevistas, métodos de observación, historias vitales, documentos, entrevistas a grupos centrados y simulaciones. Cualquiera que sea el método, o los métodos, cabe señalar que existen unos temas metodológicos referentes al trabajo de campo recurrentes que recogemos en la última parte del capítulo. Entre ellos están los temas actuales del ac-

ceso, la evaluación de grupos sensibles, la evaluación de la satisfacción de los interesados y de los usuarios del servicio, la evaluación del profesional y la evaluación participativa para la potenciación. Concluiríamos recordando que las cuestiones éticas no sólo se plantean al principio de la evaluación cualitativa, sino a lo largo de todo el trabajo de campo.

Quizá sea útil que ilustremos la conexión entre el método y la perspectiva evaluativa. Muchos evaluadores se preocupan de los aspectos *colaborativos* de su trabajo. Por ejemplo, Woods y Pollard señalan que su defensa de la síntesis de la interacción simbólica y la evaluación educativa no se basa principalmente en el deseo de impedir ningún método en particular:

Al contrario, nuestra idea tiene que ver con el propósito. Si se quiere que éste influya en la práctica del profesor y en la política de las escuelas o que las cambie, recomendamos un modelo de investigación y enseñanza alternativo, que implica trabajar *con* los agentes del cambio teniendo en cuenta sus propias orientaciones, y no trabajar *sobre* ellos poniendo en primer plano nuestros fines o propósitos (1988, pág. 15).

Objetivos similares de «trabajar con» han guiado gran parte de los trabajos en los planteamientos evaluativos en que se involucra activamente a los interesados. Weiss (1983a, 1983b), por ejemplo, defiende, aunque sea modestamente, la evaluación del interesado, ya que:

- mejora la imparcialidad de la evaluación;
- de paso, mejora la variedad de la información recogida;
- permite que los grupos locales puedan expresarse mejor, y
- democratiza el acceso a la información.

Sin embargo, también afirma que esto no se traduce en una armonía ni resuelve el conflicto entre los diversos grupos. Partiendo de estas conclusiones, el investigador podría negociar una dimensión colaborativa para una evaluación dada, pero es improbable que lo haga estrictamente sobre bases metodológicas.

Veámos en el capítulo 6 que otros abogan por modelos participativos o de potenciación y no por los simplemente colaborativos. Quienes practican la evaluación de potenciación de los evaluados normalmente desconfían de los métodos, y su deseo es «ir más allá de la

técnica». Sin embargo, los vínculos entre el propósito y los métodos no desaparecen. Vanderplaet piensa que los planes emancipadores de la evaluación pueden estar reñidos con las estrategias de evaluación disponibles. Esta falta de ajuste se manifiesta en la fragilidad de la relación entre el discurso de la potenciación y la práctica de la programación social. Aunque la autora advierte de que la preocupación singular por la técnica puede «sofocar efectivamente los conocimientos prácticos y emancipadores que se deben conseguir de esos esfuerzos», reconoce también que «la evaluación de intereses críticos no debe ignorar las cuestiones instrumentales ni privilegiarlas» (Vanderplaet, 1995, pág. 94).

Ahora queremos aplicar esta argumentación general sobre el propósito a los usos de las entrevistas, la observación, las historias vitales, los textos y documentos personales y los métodos de simulación de la evaluación cualitativa. Nuestro punto de partida es considerar la pregunta que los evaluadores cualitativos se suelen hacer respecto a los métodos: ¿Debo emplear las entrevistas, la observación o algún tipo de combinación de ambas?

EVALUAR LOS MÉTODOS DE TRABAJO DE CAMPO

Una forma de decidir los beneficios y las limitaciones relativas del uso de los métodos de entrevista o de observación en la evaluación cualitativa es comparar directamente los beneficios y las limitaciones de cada uno.

EJEMPLO 8.1. ANALIZAR LA TOMA DE DECISIONES DEL PROFESIONAL

McKeganey y sus colegas (1988) exponen una de estas comparaciones realizada como parte de un estudio de la toma de decisiones del profesional sobre la asignación de plazas en una residencia de ancianos. A los autores les interesaban los aspectos relacionados con los microprocesos de la toma de decisiones y comprender los criterios y las variaciones que intervienen en este tipo de decisiones. La tabla 8.1 resume los aspectos que se comparan y el peso que el equipo de evaluación asignó a cada uno de ellos en el contexto de su estudio.

TABLA 8.1. La entrevista y la observación para evaluar la toma de decisiones del profesional.

	Entrevistas	Observación
Nivel de los datos	Individuos	Procesos
Puntos de la decisión	Múltiples	Pocos
Triangulación de las explicaciones	Fuerte	Menos fuerte
Componentes de las decisiones	Formales	Informales
Decisiones rutinarias	Menos fuertes	Fuertes
No decisiones	Débiles	Débiles
Racionalidad / no racionalidad de las decisiones	Exagerar la racionalidad	Mucha no racionalidad
Revelación de las explicaciones privadas	Menos fuerte	Adecuada

Fuente: Basado en McKeganey y otros (1988).

La conclusión de los investigadores era que resultaba difícil emplear la observación para centrarse en las decisiones individuales, porque éstas se producen en diversos contextos. Las entrevistas, por el contrario, pueden abarcar cualquier punto de la decisión. Estos autores decidieron fijar los contextos de sus observaciones considerando los resultados de entrevistas previas. Se consideró también que la entrevista tiene mayor fuerza como medio para triangular explicaciones de diferentes profesionales. La idea de triangulación se usa mucho, a menudo de forma confusa, y nos ocuparemos de ella en el capítulo 9. De momento, observemos que la decisión de entrevistar puede mitigar el problema de sopesar las diferentes explicaciones, pero no lo resuelve.

Los entrevistadores y los entrevistados pueden tender a concentrarse sólo en los componentes formales del proceso de toma de decisiones, mientras que «uno de los beneficios del trabajo de observación es precisamente la capacidad de centrar la atención en los aspectos informales de la toma de decisiones de los profesionales» (McKeganey y otros, 1988, pág. 16). Este aspecto formal/informal se reflejaba también en la idea de los autores de que quizá sea más difícil que las personas articulen en las entrevistas las dimensiones de las decisiones que se dan por supuestas, y que se puede acceder mejor a ellas mediante la observación.

Las entrevistas pueden favorecer que se recreen decisiones antiguas como si fueran más racionales de lo que en realidad son. McKeganey y sus colegas concluyen que el trabajo de observación puede aprovechar el carácter más caótico de las decisiones actuales. Por últi-

mo, los autores creen que los profesionales pueden utilizar las categorías de la decisión privada que incluyen aspectos morales o peyorativos, quizá sobre todo cuando la demanda de un servicio sobrepasa la disponibilidad y los profesionales se ven obligados a racionarlo. La conclusión de los autores es que es menos probable que estos elementos se revelen en las entrevistas y que la observación al menos problematiza las bases de la toma de decisiones.

Después de este ejemplo, hay que hacer dos advertencias importantes. Primera, las elecciones de los métodos que se realizan como consecuencia de una comparación de este tipo no se deben tratar como absolutas, sino como contextualizadas dentro del estudio concreto. No se pueden trasladar los pros y los contras de la tabla 8.1 absolutamente a todas las evaluaciones donde se lleve a cabo una elección de métodos de este tipo. No existe ninguna razón absoluta, por ejemplo, para que los métodos de entrevista o de observación no puedan provocar no decisiones. Estos investigadores han contextualizado claramente su valoración de los dos métodos. Segunda, al comprimir las entrevistas en una opción de evaluación única y aparentemente homogénea, se omiten demasiadas cosas que son diferentes.

La entrevista evaluativa

Las entrevistas van desde la charla informal de todos los días que se da en cualquier tipo de investigación, a las entrevistas no estructuradas, que se centran en un tema concreto, las semiestructuradas, en las que se utilizan guías para la entrevista, y hasta las estructuradas, que utilizan los programas correspondientes. Por razones que derivan de la importancia del propósito local, no voy a decir que el planteamiento X de una entrevista sea generalmente el mejor para la evaluación, y que haya que evitar el planteamiento Y. Sin embargo, los enfoques de las entrevistas que se han desarrollado con la atención puesta en las aplicaciones evaluativas compensan el esfuerzo que supone traducirlas. La defensa que Holstein y Gubrium hacen de la *entrevista activa* es un ejemplo de esto; las aplicaciones de la exposición de McCracken de la *entrevista larga* semiestructurada constituyen otro nuevo ejemplo (Holstein y Gubrium, 1995; McCracken, 1988).

La tesis de Holstein y Gubrium dice más o menos lo siguiente. Las perspectivas sociológicas más antiguas tratan la conversación de la

entrevista como una cañería por la que los conocimientos pasan del entrevistado al entrevistador. Hace mucho que se dice, con cierto éxito, que en tal concepción hay un error importante: «Las dos partes que intervienen en una entrevista son necesaria e inevitablemente *activas*. Ambas desarrollan un trabajo de producción de significado» (Holstein y Gubrium, 1995, pág. 4). En los enfoques constructivista, etnográfico y posmoderno se sitúan en primer plano los procedimientos interactivos de producción de conocimientos, el *cómo* de la producción de significado. Holstein y Gubrium aprueban este planteamiento, pero piensan que la importancia capital de los «cómos» del proceso social se ha alcanzado «a expensas de los *qués* de la experiencia vivida» (1995, pág. 5). Los autores quieren readecuar la importancia de la sustancia y el contenido del proceso de construcción social.

Ya señalamos en el capítulo 3 que las versiones relativistas del constructivismo social plantean unas dificultades especiales a los investigadores de evaluación que emplean la indagación cualitativa. Parece que Holstein y Gubrium aportan un enfoque que en parte resuelve estas dificultades mediante una idea del significado como algo que no es ni predeterminado ni absolutamente exclusivo y que, por consiguiente, no se formula constantemente desde cero, sino que refleja unas condiciones locales relativamente duraderas (1995, pág. 16). Indican que la entrevista se debe estructurar como una oportunidad para la producción de relatos, en la que el entrevistado se entiende como una especie de contador de historias. Describen e ilustran esta forma de entrevistar empleando ejemplos de investigaciones sobre cuidadores que trabajan en residencias de personas que sufren demencia, sobre un organismo de control del cáncer de ámbito comunitario y sobre la calidad de vida de los residentes en centros para personas mayores.

La premisa de Holstein y Gubrium de que el objetivo de la entrevista debe ser el de propiciar la producción de relatos o narrativas tiene algunos puntos en común con el desarrollo de la entrevista cualitativa y en profundidad de McCracken, que utiliza las preguntas abiertas y directas para producir historias y relatos de casos. Miller y Crabtree la aplicaron a un estudio relacionado con la evaluación sobre cómo los médicos de familia y sus pacientes identifican e interpretan sus experiencias del dolor. Estos autores, al describir la entrevista como «un baile entre la intimidad y el distanciamiento» (Miller y Crabtree, 1992, pág. 203), subrayan la reciprocidad presente en el discurso de la entrevista y la necesidad de que el entrevistador esté abierto al dolor y a la incertidumbre que supone el autodescubrimiento.

McCracken relaciona este aspecto con el problema que se crea por el efecto de embotamiento que normalmente produce la familiaridad asentada con una cultura. Aboga por una revisión preliminar detallada de la experiencia personal relacionada con el tema en cuestión. Tal «revisión cultural» exige un examen detallado:

El investigador debe inventariar y examinar las asociaciones, los incidentes y los supuestos que rodean al tema que tiene en mente [...] La cuestión está en sacar de la propia experiencia las propiedades sistemáticas del tema (McCracken, 1988, pág. 32).

En el estudio de Miller y Crabtree sobre la interpretación y la experiencia del dolor, todos los miembros del equipo de investigación inventariaron sus propios incidentes, asociaciones y supuestos relativos al dolor que habían vivido. Los autores desarrollan la aplicación original de McCracken de dos formas, y ambas pueden ser útiles en un contexto de evaluación. En primer lugar, recomiendan que el descubrimiento de las categorías culturales se amplíe para incluir el contacto directo preliminar con los individuos pertenecientes al grupo cultural que se estudie. En segundo lugar, además del valor conceptualizador del ejercicio, sostienen que «la autoexploración también prepara una reserva de empatía para el investigador cuando el entrevistado comparte unos pensamientos y unas emociones similares» (Miller y Crabtree, 1992, pág. 199). Es claro que la incorporación de las revisiones culturales no se limita a la evaluación basada en la entrevista.

EJEMPLO 8.2. LA REVISIÓN CULTURAL

Supongamos que estoy preparando una entrevista a una mujer que tiene problemas para atender a su hijo de 25 años, que tiene unas dificultades graves de aprendizaje y vive en casa. Para realizar una revisión personal y cultural tendré que hacer dos cosas:

1. Una lista de los diversos temas en los que debería centrar mi autoindagación para una revisión cultural.
2. Una serie de notas sobre cada uno de los temas de esa lista.

Mi lista de temas podría incluir:

- ¿A quién he cuidado yo?
- Los sentimientos que en mí despertó el hecho de que algún familiar cuidara de mí.

- Mi reacción ante la realización de tareas de cuidado personal.
- Mis sentimientos sobre las fronteras entre el tiempo familiar y el tiempo laboral.
- Mi definición y uso del tiempo libre.
- Cómo entiendo el «tiempo libre» comparado con el «tiempo de trabajo».
- Las satisfacciones que produce cuidar de otros.
- ¿Cómo reacciono cuando se me limita mi tiempo?
- Atención de día y atención de noche.
- Atención en días laborables y atención en fines de semana.
- Mi equilibrio ideal entre el tiempo para estar solo y el tiempo para estar con la familia.

Después de preparar esta lista de posibles temas, la siguiente fase consiste en tomar notas sobre cada uno de ellos. Por ejemplo, he incluido los sentimientos que despertó en mí el hecho de que algún familiar cuidara de mí. Si no puedo pensar en la posibilidad de que llegue el momento en que deba enfrentarme a tener que depender físicamente de algún familiar cercano, es posible que esto provoque en mí una actitud negativa irracional hacia la persona que cuida. También he puesto en la lista mi experiencia como cuidador. En los casos en que los varones de la familia han obligado a la mujer a adoptar el papel de cuidadora, como corresponde al estereotipo, ésta será particularmente sensible al potencial discriminador que existe en la función de cuidador o cuidadora en las sociedades occidentales.

Sospecho que la entrevista plantea un problema que es más probable que aparezca en la investigación evaluativa, en la que los propios interesados internos ejercen mucha influencia, que en la investigación convencional. Como mejor se puede ilustrar es en el contexto de la evaluación relacionada con trabajos en los que la entrevista es una competencia profesional reconocida. Por ejemplo, se ha dicho en alguna ocasión que la asistencia social y algunas formas de psiquiatría son las profesiones de la entrevista por excelencia. ¿Podemos deducir de ello que estas profesiones poseen unas habilidades lo suficientemente buenas como para emplear la entrevista como un canal para la evaluación? No estoy seguro de ello. Las entrevistas se han desarrollado acertadamente como un medio para garantizar que se responda de la práctica, con el resultado de que algunos estilos de entrevista se han hecho más comunes que otros. En consecuencia, los profesionales del trabajo social tal vez no han sido lo bastante sensi-

bles a la naturaleza interactiva fundamental de toda entrevista. Hace tiempo que Cicourel (1964) decía que se debe considerar la entrevista empíricamente, como una variación de la interacción de la vida cotidiana y como un ejemplo de encuentros con extraños. Rees (1974) sugería que en la asistencia social las entrevistas se pueden entender de forma parecida. La «parcialidad» y la distorsión se deben considerar como respuestas normales y de sentido común, y se demostrará que es imposible disponer de una presentación estándar de la entrevista. La tipificación social de los roles del entrevistador, que varían en función de la cultura y la clase social, la tensión interactiva constante cuando una persona intenta penetrar en el mundo privado de otra, y las ventajas y los inconvenientes de la relación de comunicación, son dimensiones de la entrevista de las que posiblemente los trabajadores sociales no son conscientes.

Pensemos, por ejemplo, en los temas que se plantean cuando se entrevista a niños (Shaw, 1996b). La obra de Piaget y de Kohlberg ha tenido una importancia capital en el fomento de una imagen de la infancia cuyo signo distintivo es una serie de fases de desarrollo. En consecuencia, se ha pensado que el trabajo cualitativo con niños consiste en descubrir de qué forma el «que sabe» puede descubrir cosas sobre el niño, teniendo en cuenta la edad de éste. Se piensa que la especial naturaleza de la entrevista con niños deriva de la limitada capacidad de éstos para comprender el lenguaje, exponer una experiencia subjetiva y mantener unas determinadas relaciones afectivas (Parker, 1984). Este modelo de carencias se manifiesta cuando los investigadores advierten a sus lectores sobre este tipo de entrevistas. Por ejemplo, Bierman y Schwartz, al hablar de las entrevistas clínicas, advierten de que los niños:

tienen problemas para tener en la mente más de un concepto a la vez. Además, sus concepciones *se limitan* a ideas que puedan adoptar formas concretas y observables. Son *incapaces* de considerar perspectivas alternativas [...] *Simplemente aprenden* a interpretar las emociones de los demás (1986, pág. 268; la cursiva es mía).

Esta idea de la infancia como unas fases evolutivas que se van desplegando ha limitado los métodos de muchos investigadores y ha conducido a un modelo de entrevista infantil que se piensa que «puede ayudar al clínico a valorar la complejidad evolutiva del estilo interpersonal del niño y su razonamiento social y emotivo» (1986, pág. 276).

Un modelo de carencias de este tipo tiene unas consecuencias preocupantes para la forma de entender las culturas de los niños. En efec-

to, no es probable que se interprete que éstos tienen su propia cultura o, en el mejor de los casos, su cultura se entenderá según su aproximación a la cultura del mundo de los adultos. Una de las principales fuerzas que inició un esperado correctivo de este tipo de investigación fue la obra de Cicourel sobre el lenguaje de los niños, que presentaba la idea de éstos como capaces de tener su propia cultura, que ellos mismos transmiten de una generación a otra. Desde esta perspectiva tan diferente, se considera que los niños poseen una competencia interpretativa igual, o incluso mayor, que la de los adultos y potencialmente son «más creativos, más honestos, en resumen, en cierto modo más humanos que las personas mayores» (Goode, 1986, pág. 84). Para Goode, el modelo del niño como una *tabula rasa*, que crece siguiendo unas fases evolutivas, supone un «adulterismo» etnocéntrico en el que los investigadores demuestran ser unos padres que escriben unas versiones un tanto abstractas de sus propios hijos y de los hijos de otros.

Las estrategias de la entrevista en la investigación con niños y jóvenes se han desarrollado desde esta posición en dos direcciones. En primer lugar, se ha aceptado gustosamente y de forma bastante generalizada la idea de los niños como personas que poseen unos derechos similares a los de los adultos en lo que se refiere a otorgar el consentimiento informado para llevar a cabo una investigación. Por ejemplo, Butler y Williamson (1994) ofrecían a los niños la posibilidad de decidir cómo deseaban que se les entrevistara, de forma individual, en parejas o en grupos. Realizaban sesiones de preparación con ellos para transmitir este espíritu y la libertad de seguir una determinada línea de razonamiento. Dejaban que ellos mismos controlaran la grabadora y la duración de la entrevista. En los casos de entrevistas individuales, aceptaban que éstas «no se prestan a estructuras rígidas y a secuencias definidas. Los niños van dando saltos, y los investigadores deben saltar con ellos» (Butler y Williamson, 1994, pág. 30). Esto solía incluir el reaccionar al humor de los niños y reconocer que «atienden con seriedad aunque se oculten tras un caparazón de humor» (1994, pág. 46).

El segundo modelo de entrevista con niños y jóvenes tiene un mayor alcance. Básicamente implica un cambio importante en la forma de entender la entrevista, de manera que ya no se considera una fuente de información proporcionada por los jóvenes informantes, sino como unos datos en sí misma. Por ejemplo, hablar con adolescentes no constituye un medio para conseguir información sobre los procesos de socialización sino que, desde esta perspectiva, es en sí mismo

un caso de socialización. La «segunda mirada» de las entrevistas a los adolescentes de la que habla Baker ejemplifica este enfoque. Dice la autora que «la clave para convertir la entrevista en un recurso es entenderla como un suceso real del mundo real, y no sólo como un medio de acceso a unos datos, sino como un dato en sí misma» (Baker, 1983, págs. 501, 508). Esto le permite entender el carácter de las entrevistas como encuentros sociales entre adultos y jóvenes, en los que los primeros hacen preguntas a los segundos. Las preguntas contienen una teoría implícita adulta sobre la adolescencia, y las respuestas contienen una teoría implícita adolescente sobre la madurez. Las entrevistas son una ocasión para la gestión de la identidad:

Las entrevistas son enclaves de la socialización. En estas entrevistas se invita a los adolescentes a mostrar su pertenencia a la comunidad del lenguaje de los adultos, al interpretar las preguntas y desarrollar un discurso con los entrevistadores (Baker, 1983, págs. 516-517).

Siebert concluye asimismo que «los datos de las entrevistas que realiza un entrevistador adulto permiten comprender mejor las competencias que los niños muestran en la interacción con las personas mayores [...] Sin embargo, es cuestionable si estas conclusiones se pueden generalizar a otros contextos, especialmente a los contextos de interacción entre iguales» (Siebert, 1986, pág. 373). Las conclusiones de este tipo tienen unas implicaciones evidentes para el encuentro en el trabajo de campo de la evaluación que implique una charla entre adultos y niños en los ámbitos de la salud, la asistencia social y la educación.

La observación

Los métodos de la observación han de estar mediatizados también por el propósito de la evaluación (Tanner y Le Riche, 1999). Estar con niños y jóvenes y comprender el mundo de la infancia es a la vez algo fascinante y misterioso. ¿Cuántos de nosotros hemos hurgado en los libros de Iona y Peter Opie para aprender y recordar (Opie y Opie, 1959, 1969)? El problema de la «Sociedad de los Chavales» (la expresión es de Glassner) es su cercanía física y su distancia social: «Nuestra proximidad a los niños nos puede llevar a pensar que estamos más cerca de ellos de lo que en realidad estamos, que la única diferencia es que [...] los niños están creciendo aún [...] y que a me-

nudo se equivocan» (Fine y Sandstrom, 1988, pág. 34). El valor de la observación del participante es que debemos convertir a estos vecinos en extraños y, por ese camino, en iguales, si queremos entender qué significa ser niño y si deseamos ser capaces de ver el mundo «a través de sus corazones y sus mentes» (1988, pág. 12). De este modo batallaremos por excavar en nuestra propia experiencia y nos enfrentaremos al desafío de abrir una brecha en nuestras sólidas defensas.

Se han desarrollado tres tipos de estudios observacionales de niños. Están las observaciones en que se descubre que los niños son más maduros y capaces de lo que pensamos; las observaciones que desromantizan la infancia y en las que se descubre que los niños son tendenciosos y rebeldes; y, por último, las investigaciones que aclaran cómo se produce el proceso secreto de la educación o de la socialización mediante iguales. Fine y Sandstrom destacan que «no conocemos ningún estudio que haya descubierto que los niños son más “infantiles” de lo que les hemos reconocido» (1988, pág. 72).

El reto que se les plantea a los evaluadores que deciden realizar observaciones con niños es «captar la dinámica de las interacciones de los niños y encajar en sus actos interpretativos sin perturbar el flujo» (Mandell, 1988, pág. 464). Tal vez la decisión más importante que el observador habrá de tomar al decidir observar a niños es definir la naturaleza del rol de observador, que irá desde la función del observador que se mantiene al margen, al rol de «amigo especial» que adopta Corsaro (1985) en su excelente estudio de los niños de una guardería, hasta el «rol menos adulto» activo y totalmente participativo que practicaba Mandell (1988) en su observación de centros de preescolar. Mandell aboga por dejar en suspenso todas las características del adulto, a excepción del tamaño físico, y adoptar un papel que reduzca al mínimo la distancia social, suspenda los juicios sobre la inmadurez de los niños, y suponga implicarse en una acción conjunta con éstos y correr el riesgo de que a uno le tomen por loco.

Las historias de la vida

En el trabajo sobre las historias de la vida hay una serie de desarrollos que tienen su valor para el trabajo de campo de la evaluación cualitativa. El estudio de la historia de la vida es un terreno en el que a menudo han corrido paralelos la etnografía y los métodos clínicos (Martin, 1995, 1999). Como dice Cohler:

El concepto de narrativa, entendido como [...] un pasado recordado hoy, un presente experimentado y un futuro anticipado, es básico tanto en el estudio sistemático a lo largo del tiempo como en el esfuerzo por reducir la experiencia de desasosiego personal (1994, pág. 167).

La proximidad queda bien ilustrada en la exposición que Bowen hace en su artículo «Las delicias de aprender a aplicar el método de la historia de la vida a quienes no asisten a la escuela» (Bowen, 1993). El autor tuvo unas doce reuniones con cada una de las cuatro personas que habían tenido graves problemas educativos en ese momento, o por entonces, y concluyó con ocho o nueve horas de conversaciones grabadas. Tres de estas personas escribieron unas autobiografías de entre 5.000 y 12.000 palabras.

La evaluación de potenciación ha propiciado el desarrollo de las historias de la vida críticas (Barone, 1992; Clifford, 1994). Barone describe los beneficios que en su opinión proporciona la narración crítica de historias. Los autores de esos relatos «tratan de hacer palpables y comprensibles la pena y la crueldad del aislamiento a que se ven sometidas las personas» (Barone, 1992, pág. 146). Describe las fases por las que avanza tal «resolución cualitativa de los problemas» y pone un ejemplo de un estudio de casos de Billy Charles Barnett, que fue un fracasado escolar en potencia.

- Un catálogo de fragmentos incoherentes a medida que «los fenómenos parpadean sin ajustarse a un patrón discernible» (1992, pág. 144).
- Una relación provisional con «fragmentos estructurados» (1992, pág. 144).
- Surge una metáfora o una cualidad omnipresente. En el caso de Billy Barnett fue la idea de riesgo y de combatir la civilización permanentemente cercenante de la que la escuela formaba parte. Billy se convierte en un símbolo para todos los acosados por «el peligro de las instituciones educativas» (1992, pág. 145).
- La selección del material y la historia final.

Barone concluye: «Confío en que cada vez sean más los investigadores cualitativos actuales que empleen sus privilegios para contar historias con las que los lectores puedan situar el origen de ese dolor» (1992, pág. 146).

Un enfoque distinto es el que representa el desarrollo del interés por la «sociología del curso de la vida». Una vez más, hay que insistir

en que este tipo de métodos no son intrínsecamente una «evaluación» y, por lo general, los han empleado investigadores que no considerarían que su trabajo es una evaluación. Es un requisito permanente que estos métodos se cultiven en función del propósito evaluativo y que se ajusten a él. El análisis del curso de la vida «sitúa el cambio y el desarrollo en el centro del análisis, y se propone explorar la dimensión temporal inevitable de nuestras vidas y experiencias» (Morgan, 1985, pág. 177).

Compartir una historia reporta unos beneficios evidentes, entre ellos el sentido de la perspectiva, el autoconocimiento, la creación de una comunidad y un mejor sentido de cómo quiere uno que termine su propia historia (Atkinson, 1998). Sin embargo, Atkinson corre el riesgo de sentimentalizar el proceso de narración de historias. Selwyn, por ejemplo, registra un estudio en que unos posibles padres adoptivos le contaron sus historias de cómo llegaron a solicitar la adopción, y su reacción cuando se les negó lo que solicitaban. Las entrevistas no sólo muestran su dolorosa experiencia del rechazo, sino que en algunos casos sirvieron también para despertar de nuevo unos sentimientos en parte contenidos de enojo, de inutilidad, de vacío y de un futuro de soledad (Selwyn, 1991, 1996).

EJEMPLO 8.3. PROBLEMAS CON LA VIVIENDA

Clapham, Means y Munro señalan que las investigaciones sobre las asignaciones de vivienda o alojamiento han tendido a ofrecer imágenes instantáneas, sin dar una idea de cómo han llegado las personas a su situación actual en lo que a la vivienda se refiere. Defienden que la vejez y la vivienda en esta edad se entiendan como

fases del curso de la vida que sólo se pueden entender haciendo referencia a la experiencia previa y a las actitudes y las decisiones de las personas, y a las oportunidades que se les abrieron en el transcurso de toda su vida (Clapham y otros, 1993, pág. 133).

Los investigadores proceden mediante la construcción de una biografía personal a través de entrevistas cualitativas exhaustivas, para comprender cómo llegaron las personas a la situación actual en lo que a su situación respecto a la vivienda se refiere. El objetivo de la entrevista es descubrir por qué se tomaron determinadas decisiones y no se tomaron otras. El sistema incorpora el reconocimiento de que las diferencias de clase, sexo y etnia no desaparecen con la edad. A diferen-

cia de un enfoque puramente biográfico, aquí se distingue entre los diferentes tiempos en que se estructura la situación de la vivienda (la edad, el estado de salud, etc.); el tiempo familiar (la fase del ciclo del tiempo familiar en que se tomaron o se impusieron determinadas decisiones); y el tiempo histórico (las condiciones económicas, sociales y políticas dominantes).

Textos y documentos

La experiencia, las historias y los textos de alumnos, padres, pacientes, cuidadores y usuarios del servicio, y su nueva narración en los relatos informales que tienen lugar en la conversación con el profesional, los registros de casos, los informes, etc., constituyen una reserva rica y a veces enigmática de material relevante para la evaluación. Muchas de las orientaciones que se dan en la literatura se pueden leer y asumir sin necesidad de reelaborarlas mucho. Entre ellas está el debate sobre las distinciones entre los textos orales y escritos, la importancia de los públicos para los que en principio se concibieron los textos en cuestión, los niveles variantes de acceso a esos textos y los criterios de autenticidad, exactitud, sinceridad, representatividad y significado con los que se pueden valorar (Scott, 1990).

La mayor parte de todo esto se ha asumido en la literatura sobre metodología de evaluación y en los informes sobre estudios evaluativos. Pero existen algunas lagunas en el uso de los registros en la evaluación. Por ejemplo, hay una dependencia casi exclusiva de los documentos oficiales y cuasi-oficiales, normalmente los que se generan en el programa que patrocina la organización en cuestión, y en la evaluación del programa que ésta facilita. Los evaluadores cualitativos deberán ser sensibles a las oportunidades de utilizar otras formas de texto, en especial en la evaluación que muestre los objetivos de los interesados, en la evaluación del profesional y en la evaluación participativa o de potenciación.

Clandinin y Connelly (1994) han realizado una revisión precisa de los métodos de la experiencia personal en la investigación cualitativa, partiendo de su trabajo de investigación previo llevado a cabo en el ámbito escolar. Distinguen entre anales y crónicas. Los *anales* se pueden entender como una «línea» que esquematiza la vida de una persona, que los sucesos, los lugares, los años o los recuerdos importantes dividen en momentos o segmentos. De este modo es posible tener

una sensación de conjunto, incluidos los altibajos y los ritmos que las personas construyen en torno a su ciclo vital. Las *crónicas* perfeccionan algún punto de los anales. Unos y otras son, en su conjunto, «una forma de asentar sus historias orales» (1994, pág. 420), como si de un andamio se tratara, y de reunir sus experiencias. También son una forma de empezar a oír las historias familiares de la persona. Estas historias sobre los miembros y los acontecimientos familiares se transmiten de generación en generación. Con ellas las personas descubren su propia identidad, tanto internamente como en la relación de la familia con el mundo. Las fotografías, la correspondencia, los objetos y demás cosas marcan los momentos, los acontecimientos y a las personas y, en torno a todo ello, se construyen y se reconstruyen las historias, a menudo por parte de las mujeres, que las cuentan a los hijos.

Los diarios sirven para explicar la experiencia. Clandinin y Connelly hablan de una hermosa analogía y comparan el efecto que las entradas de un diario producían en una mujer con unos caramelos diminutos que, separados, no tienen sabor alguno, pero juntos son toda una golosina. Es frecuente que los niños y los adolescentes lleven su diario, en el que anotan lo que piensan, lo que hacen y sus historias. La mayoría de esos diarios se mantienen privados, y sólo algún accidente de la historia los saca a la luz pública, como en el caso de Ana Frank o de las hermanas Brönte. Los textos en que se plasman recuerdos infantiles suelen reflejar emociones recogidas en el sosiego que proporciona la madurez. Pero es posible, por ejemplo, que algunos niños y jóvenes escriban sus diarios como un intento de comprender sus experiencias, «captar los fragmentos de experiencia con la intención de entenderlos» (Clandinin y Connelly, 1994, pág. 421). Cuando los adultos hablan de las experiencias de su niñez, a quien escucha se le plantean unas dificultades especiales. ¿Quién es el que habla?

¿Es el adulto que interpreta la experiencia de la infancia, en cuyo caso es el adulto quien habla? ¿O es el adulto que expresa la historia del niño tal como éste la hubiera contado, en cuyo caso el que habla es el niño? (1994, pág. 424).

Los textos en que se expone la experiencia personal revelan la escena y la trama, las dimensiones del lugar y del tiempo. La trama, el significado y la interpretación distan de ser algo constante. Los significados de hoy pueden convertirse en elementos de una crónica de sucesos de mañana, a medida que los participantes cambian su forma de interpretar las cosas.

Los grupos centrados

Uno de los avances más prometedores de la aplicación de la evaluación cualitativa y participativa se ha producido con los trabajos sobre los grupos centrados (Dean, 1994; Kitzinger, 1994; Kreuger, 1994; Morgan, 1993, 1997). Los grupos centrados tienen su origen en la investigación de mercado, aunque también recibieron su impulso en la obra del sociólogo estadounidense Robert Merton sobre las entrevistas centradas. Éstas consisten en charlas en grupo organizadas para explorar una serie concreta de cuestiones: «El grupo está *centrado* en el sentido de que implica cierto tipo de actividad colectiva, por ejemplo ver una película, evaluar un mensaje concreto de contenido sanitario, o simplemente hablar de una determinada serie de cosas» (Kitzinger, 1994, pág. 159). La planificación y el diseño de los grupos centrados se basan sobre todo en las decisiones sobre el papel del moderador del grupo, el desarrollo de los planes del grupo y las decisiones sobre la composición de este último. En la actualidad, se trabaja con grupos homogéneos: «Se cree que tener sesiones separadas con grupos homogéneos aunque opuestos produce una información más exhaustiva que la que se conseguiría con grupos heterogéneos» (Knodel, 1993, pág. 40).

Los grupos centrados tienen tres ventajas particulares para la evaluación cualitativa. En primer lugar, la propia interacción del grupo constituye unos datos —«la interacción es el método» (Jarret, 1993, pág. 198)—. Kitzinger dice en resumen que el método «permite que el investigador pueda examinar los diferentes puntos de vista de las personas mientras operan en una red social, y estudiar cómo se construyen, se expresan, se censuran, se rebaten y se cambian las explicaciones a través de la interacción social» (1994, pág. 159).

En segundo lugar, los grupos centrados son una forma de evaluación participativa. Tienen valor cuando hay una diferencia de poder entre los participantes y quienes toman las decisiones y, por consiguiente, tienen un potencial considerable en la aplicación dentro de la evaluación cualitativa. Por último, suponen un valioso enfoque para descubrir el grado de consenso sobre un tema particular: «Los coparticipantes actúan como coinvestigadores, llevando así la investigación en direcciones nuevas y muchas veces inesperadas, y participando todos de forma a la vez complementaria [...] y argumentativa» (Kitzinger, 1994, pág. 166).

La mayoría de los autores que se ocupan de este campo se oponen a la aplicación de los grupos centrados a nada que no sea los fi-

nes de la investigación o de la evaluación formal. En mi opinión, se trata de una generalización inútil. Los grupos centrados también pueden hacer su particular aportación a la evaluación participativa y del profesional. Los ejercicios de formulación de problemas, desarrollo de proyectos, prácticas antidiscriminatorias, tratamiento de proyectos que se han quedado «estancados», ejercicios de retroalimentación del consumidor, y la evaluación de temas sensibles... en todo ello los grupos centrados pueden ofrecer algo a la evaluación cualitativa. Gibbs resume los argumentos en favor de los grupos centrados:

Son especialmente útiles cuando existen unas diferencias de poder entre los participantes y los profesionales o quienes toman las decisiones, cuando interesa el uso cotidiano del lenguaje y la cultura de determinados grupos, y cuando se quiere estudiar el grado de consenso sobre un determinado tema (1997).

Pese a lo prometedor de este método, los grupos centrados se deben utilizar dentro de sus limitaciones. La actual adopción casi instintiva de las entrevistas de grupo denominadas grupos centrados, sea cual sea su modo de funcionamiento, supone toda una confusión. Los grupos centrados, por ejemplo, siempre dependen de la interacción dentro del grupo, basada en los temas que proporciona el investigador. En los estudios hay pocas pruebas, o ninguna, de los beneficios de los grupos centrados frente a los métodos de la entrevista (pero véase Kitzinger), y hay situaciones en las que probablemente no habría que usar los grupos centrados. No convendría emplearlos si la intención del profesional es mejorar la comunicación del participante o las habilidades del grupo. En un sentido más general, no se deben adoptar con fines terapéuticos, ni cuando el objetivo principal sea garantizar la acción inmediata. Si la información, la comprensión o la explicación no son esenciales en los planes del grupo, entonces se deben emplear otros métodos de intervención. También puede resultar difícil que el investigador identifique con claridad los mensajes individuales dentro del grupo. Si las opiniones personales no se pueden expresar perfectamente en ese contexto, o si es previsible que no se respete la confidencialidad, no hay que utilizar el método. Cuando los miembros del grupo se conocen entre sí particularmente bien, tampoco son aconsejables los grupos centrados. Por último, si se tienen en cuenta los beneficios que reporta pertenecer a un grupo homogéneo, no es recomendable formar grupos centrados

en los que participen tanto los pacientes, los alumnos y los cuidadores como los evaluadores y los profesionales.

Las simulaciones

Las simulaciones tienen el potencial de proporcionar «una herramienta exclusiva e innovadora que aún no se ha aplicado de forma amplia» en la evaluación (Turner y Zimmerman, 1994, pág. 335). Tienen dos aplicaciones principales: primera, como prueba evaluativa de la discriminación de un servicio y, segunda, como representante cualitativo del control dentro de un enclave natural.

Lo habitual es que la primera aplicación conlleve representar pares de roles (hombre/mujer; blanco/negro; no discapacitado/discapacitado). Las parejas se presentan por separado a un determinado servicio y como similares desde el punto de vista de las necesidades y de los derechos al servicio. «Hasta la fecha, la representación de pares de roles se ha empleado principalmente para comprobar la discriminación racial y étnica en la asignación de viviendas y de trabajo» (Turner y Zimmerman, 1994, pág. 332), pero el método funcionaría también para evaluar el reparto de algún recurso escaso, donde intervienen el juicio y la discreción. Por ejemplo, se ha preparado una variante del método en el que se utiliza a un solo investigador, y en el que un estudiante de graduado social profesionalmente cualificado y que padece una pequeña discapacidad visual solicita unos puestos de trabajador social que se anuncian; se cambia de método sistemáticamente en el caso de que se exponga la discapacidad, para comprobar el funcionamiento de la política de empleo no discriminatoria.

El desarrollo de la segunda aplicación es más reciente, y representa un avance en el uso de las viñetas en la investigación de políticas. Quienes evalúan el proceso de la práctica profesional se enfrentan a la invisibilidad de la práctica. ¿Cómo se pueden descubrir las formas que adopta la práctica de los abogados, los contables, los profesionales de la medicina o los asistentes sociales? ¿Cómo tratarían el mismo caso unos profesionales distintos? Los métodos de los que hemos hablado hasta ahora no abordan directamente este problema. No obstante, Wasoff y Dobash (1992, 1996) emplearon un método innovador y prometedor en la práctica de los abogados municipales. El uso de clientes simulados en situaciones «naturales» les permitió identificar unas variaciones en la práctica que se podían atribuir con cier-

ta seguridad a las diferencias entre los abogados, más que a los artefactos de las diferencias entre los casos.

Supongamos, por ejemplo, que uno desea realizar una evaluación cualitativa de las decisiones que toman los responsables de la vivienda, el personal médico y los asistentes sociales, respecto a la asignación de paquetes de gestión de cuidados. Los evaluadores que usen clientes simulados prepararán un pequeño número de historias de casos detalladas, diseñadas para comprobar las decisiones que se someten a consideración. Un investigador o un evaluador representa el papel del cliente en la historia. El gestor de la vivienda, el personal médico y los trabajadores sociales entrevistan al «cliente» en el entorno «natural» de su trabajo.

El método es un desarrollo de los métodos de simulación de roles, pero es un tanto diferente. La simulación se ha empleado mucho en los cursos de formación, y mucho menos en la práctica de la evaluación. Además, se ha usado para desarrollar destrezas y no para comprender ni para evaluar. La puesta en práctica del ejercicio es también muy diferente de la simulación de papeles, en el sentido de que el escenario es mucho más detallado, y el investigador que actúa de cliente o paciente es el único que representa un papel. El profesional desempeña su papel real y el enclave o la situación no se construyen.

El uso de clientes simulados tiene varios elementos a su favor. En primer lugar, es probable que a algunos evaluadores les resulte conocida la «familia» de métodos de la que procede. En segundo lugar, otros métodos no siempre son viables, por razones prácticas o éticas. Los clientes simulados evitan los problemas éticos que supone tener que buscar la cooperación de clientes auténticos. En tercer lugar, y sobre todo, hace visible la práctica. Con esta breve descripción se verá claramente que el método no podría ser una herramienta para evaluar casos concretos, sino que se centraría en *tipos* de práctica específicos.

La aplicación de los métodos de simulación tiene sus limitaciones. El método requiere unos recursos adicionales para preparar el material del caso, quizá para representar el papel de clientes y para reflexionar sobre el material bastante detallado que se obtiene de las transcripciones de las entrevistas. Por consiguiente, es previsible que el coste sea relativamente elevado y exige unos niveles razonablemente altos de habilidad para la investigación. Es mucho más difícil utilizar el método para comprobar la discriminación en otros niveles de las organizaciones, debido a los temas éticos adicionales que plantea la evaluación encubierta prolongada. Por último, el uso de la si-

mulación para comprobar la discriminación corre el riesgo de suscitar preocupaciones por la posibilidad de que se pueda incitar a la comisión de infracciones.

TEMAS Y CUESTIONES DEL TRABAJO DE CAMPO DE LA EVALUACIÓN CUALITATIVA

La elección de unos métodos específicos no se puede considerar ajena a los temas y las cuestiones de que hablábamos anteriormente. Entre ellos están los temas del acceso inicial y continuado, la frecuencia con que la evaluación supone estar en contacto con grupos en los que la investigación plantea cuestiones sensibles, y los problemas que se suscitan al calcular el grado de satisfacción del alumno, el paciente, el cliente o el usuario del servicio. Los temas del trabajo de campo están presentes también en la evaluación del profesional y en la evaluación participativa y de potenciación. Por último, en toda la fase de recogida de datos de la investigación evaluativa se plantean problemas éticos. En los párrafos que siguen, señalaremos los aspectos principales de todas estas cuestiones.

Yin recomienda también la creación de una base de datos del estudio de casos, en la que se incluyan las notas de campo, los documentos y las narrativas o relatos, y para mantener una cadena de las pruebas: «El principio es permitir al observador ajeno [...] seguir el curso de cualquier prueba, desde las preguntas de investigación iniciales hasta las conclusiones definitivas del estudio» (Yin, 1994, pág. 98).

El acceso

El acceso no es simplemente un conjunto herméticamente sellado de negociaciones que tienen lugar en la fase de diseño de la investigación evaluativa: «Las combinaciones posibles de negociador, enclave y atributos del participante generan una letanía casi interminable de “situaciones de acceso” distintas» (Lofland y Lofland, 1995, pág. 22). Gran parte de la experiencia y de las orientaciones desarrolladas en la literatura sobre investigación tradicional se aplica perfectamente a la evaluación, y no hace falta que aquí las repitamos (para una exposición práctica, véase Hornsby-Smith, 1993).

La relación del evaluador con el enclave quizá sea el problema que con mayor frecuencia se interpreta mal en la investigación evaluati-

va. La razón está en la tendencia recurrente de los evaluadores a sobrestimar el valor de iniciar la evaluación como alguien perteneciente a ese enclave. No se puede decir lo mismo de todos los teóricos. Veámos en las primeras páginas del libro hasta qué punto Scriven (1997) condena la investigación desde dentro. Hay buenas razones para ser cautos. El estatus de persona de la casa en realidad puede *limitar* el acceso, ya que, a pesar de que el estatus de que ya goza esa persona de la casa es probable que dé acceso a algunas partes de la organización, la misma probabilidad existe de que impida el acceso a otras. Los problemas de dar teóricamente por supuestas las cosas son más evidentes para las personas de la propia organización, que es más probable que piensen que ya «saben» cómo son realmente las cosas y, en consecuencia, entorpezcan el potencial de la extrañeza antropológica del que llega de fuera. Refiriéndose a los profesores como investigadores, Burgess dice:

A muchos profesores les ha sido difícil cuestionar conceptos tan cotidianos como los de enseñanza, aprendizaje y evaluación, y las definiciones de alumnos «buenos» y «malos». Además, varios profesores han tenido dificultad para plantear preguntas simples sobre las metas y los objetivos de un centro escolar (1980, pág. 169).

Los Lofland concluyen:

El principio es: no desesperar ni confiar en exceso en lo que se refiere a la relación con el enclave. Cualquiera que sea la relación, es a la vez una ventaja y un inconveniente (Lofland y Lofland, 1995, pág. 23).

El tema casi siempre asoma en la evaluación del profesional. Por ejemplo, el profesor investigador se enfrenta al dilema de si conviene emprender o no una evaluación encubierta, con los consiguientes problemas de tener un acceso limitado. Otra posibilidad es que realice una investigación semiencubierta, de la que informaría a sus colegas y a sus alumnos, pero no al director del centro. O al revés, hablaría de la investigación con el director, pero con nadie más. Burgess piensa que esta última opción es la estrategia más frecuente. Sin embargo, plantea el problema añadido de que, aunque aporte un apoyo, el profesor puede haber representado el papel del espía o el colaborador. «Es posible que los alumnos vean en el profesor investigador más al profesor espía que al investigador» (Burgess, 1980, pág. 168). Los dilemas éticos consiguientes son reales: «En estudios realizados

sobre los deberes para casa, los profesores investigadores han encontrado pruebas de que hay profesores que no ponen este tipo de deberes, y de alumnos que les dedican poco tiempo» (1980, pág. 168).

Davies y Kelly (1976) recordaban hace unos años, con sensibilidad y honradez, en qué grado los problemas de acceso, y las preocupaciones éticas y políticas consiguientes, pueden cambiar el panorama durante una evaluación. Hacen una exposición retrospectiva de una evaluación de observación de participantes en un proyecto municipal del norte de Inglaterra, dedicado a jóvenes sin hogar ni futuro. Los Lofland (1995) señalan que el éxito de negociación entre partes parecidas a las que se encontraron Davies y Kelly depende:

- De que los participantes perciban que el investigador es realmente alguien ajeno al conflicto.
- De la percepción permanente de la imparcialidad del investigador respecto al conflicto.
- De que el investigador demuestre una honradez escrupulosa y estudie ambas partes, además de mantener con escrúpulo el anonimato en lo que se refiere a los contenidos de las ideas y las estrategias privadas de cada parte (es decir, no actuar de agente doble).
- De poseer una buena cantidad de información exacta sobre lo que sabe cada parte de la otra mientras dure el conflicto.

Evaluación y temas delicados

En este capítulo he empleado deliberadamente muchos ejemplos de evaluaciones en las que intervenían niños y jóvenes, en parte para señalar que los evaluadores cualitativos deben practicar un tipo de investigación que plantea muchos temas dedicados, una delicadeza que puede ser inherente al problema evaluativo, a la metodología, o a la relación entre los intereses de los participantes en la indagación. Evidentemente, no pensamos que podamos decidir *a priori* si una evaluación va a ser «delicada» o no. *Cualquier* evaluación variará en su grado de «delicadeza», sea en sus distintos participantes o en sus diferentes fases.

La evaluación no plantea problemas de delicadeza que sean materialmente diferentes de aquellos que se plantean en la investigación cualitativa, sobre lo cual hay una literatura cada vez más abundante (Lee, 1993; Renzetti y Lee, 1992; Zeller, 1993), una literatura que, la-

mentablemente, ha de encontrar aún su sitio en la literatura sobre evaluación. McKeganey y Bloor (1991) destacan un ejemplo olvidado cuando hablan de que las relaciones entre los participantes plantean problemas de género. Al reflexionar sobre la evaluación participativa en comunidades terapéuticas, en una unidad para drogadictos jóvenes y en los servicios de ayuda en el hogar, los autores concluían que hay dos de esos problemas de género en las relaciones a los que probablemente no se va a prestar la atención suficiente. En primer lugar, aunque la ceguera cultural ante las relaciones entre los del mismo sexo hacía que el acceso resultara más fácil a los varones como investigadores masculinos, las actividades y las conversaciones en las relaciones dentro de un mismo sexo seguían mostrando un acusado sexismo. En segundo lugar, en las relaciones entre los dos sexos se pueden hacer interpretaciones sexuales de las que el investigador no sea consciente:

En la vida cotidiana se produce un deslizamiento recurrente entre las relaciones entre sexos y las relaciones con un componente sexual. Muchas de las relaciones asexuales de todos los días a menudo se convierten en problemáticas o ambiguas para uno o ambos participantes, cuando una de las partes considera una posible interpretación sexual de algo que la otra dice o hace (McKeganey y Bloor, 1991, pág. 209).

La satisfacción

Los chicos habían dicho que les había divertido y probablemente dirían que les hizo pensar. Probablemente tenían menos fe que yo en su efectividad. Algunos se mostraban un tanto críticos. «Sí, está bien que te puedas sentar aquí a hablar de todo esto, pero en el mundo real las cosas no son así.»

Funcionario encargado del seguimiento de personas
en libertad condicional

Este funcionario inglés refleja uno de los problemas que se suscitan cuando se pregunta por la satisfacción. Existen pruebas abundantes de que las encuestas sobre satisfacción producen una proporción uniformemente elevada de respuestas positivas. En consecuencia, normalmente no ha sido viable poder diferenciar qué usuarios de un servicio tienen más o menos probabilidades de sentirse satisfechos, y las encuestas sobre satisfacción muchas veces no dicen qué hay que

cambiar. Una estrategia que a primera vista parece que puede potenciar a los usuarios del servicio, después resulta que fomenta el conservadurismo y la inacción. Se crea el peligro de que «a través de unos elevados niveles de expresiones de satisfacción que relativamente carecen de sentido, se crea una ilusión de protección al consumidor, que casi siempre no hace más que reafirmar el *statu quo*» (Williams, 1994, pág. 515). Para entender por qué ocurre así, debemos analizar qué pasa cuando se pide a los consumidores que evalúen la prestación de un servicio.

Por ejemplo, los ciudadanos pueden pensar que no tienen control alguno sobre los resultados de su colaboración con los funcionarios públicos. Sin esta medida de influencia, la propia idea de evaluación probablemente parecerá extraña y carente de sentido. Además, si un servicio se considera más como un favor que como una prestación a la que la persona tiene derecho, o como una sanción (por una conducta delictiva o un incumplimiento personal), entonces la crítica se entenderá como una estrategia muy arriesgada, sin ningún beneficio evidente. Los defensores de la protección del consumidor ignoran con demasiada facilidad las diferencias de poder que existen entre los profesionales y sus presuntos beneficiarios.

Cuanto más impotentes se consideran las personas, más probabilidades hay de que ajusten sus aspiraciones referentes a lo que razonablemente pueden esperar conseguir del intercambio. Pueden llegar a adoptar lo que Fisher y sus colegas, en su trabajo sobre las experiencias de los niños que están bajo la tutela de la autoridad local, reconocían como una fuerte sensación de resignación fatalista, una aceptación pasiva, especialmente por parte de las madres, de la pérdida de los derechos paternos sobre sus hijos (Fisher, Marsh y Phillips, 1986). De modo similar, el papel de enfermos que representan los pacientes de un hospital exige una pasividad por su parte. Williams, en su útil disquisición sobre si la satisfacción es un concepto válido, destaca que esta pasividad puede socavar la base de la propia idea de legitimidad de la evaluación: «En tal situación, se podría decir que la satisfacción del paciente es básicamente el reflejo del rol que los pacientes adoptan en relación con los profesionales de la salud, con independencia de la calidad de la atención» (1994, pág. 513). Debido a esta relación de dependencia, los consumidores se pueden sentir a la vez agradecidos e insatisfechos por un servicio y la calidad de la atención que se presta puede ser peor de lo que parecen indicar las encuestas sobre satisfacción. Dicho en otras palabras, «la protección del consumidor depende de la negativa a aceptar el paternalismo;

se basa en la existencia de consumidores, y no de pacientes pasivos» (1994, pág. 514).

Un último problema surge de la apreciación de que las evaluaciones del consumidor «son relativas al contexto, al conocimiento de los servicios disponibles, a las expectativas, a la ayuda recibida en encuentros anteriores, a la ayuda recibida de otras fuentes, a las percepciones de la “simpatía” del asistente social» (Rees y Wallace, 1982, pág. 72). La cuestión no es en modo alguno exclusiva de la asistencia social. Una expresión de satisfacción ante un determinado servicio es un juicio relativo, una comparación entre las aspiraciones y el estatus y bienestar que se perciben después de recibir el servicio. Las mediciones descontextualizadas de la satisfacción no son suficientes. Los evaluadores y los profesionales también necesitan conocer la naturaleza y el nivel de las aspiraciones del usuario y su estatus y bienestar autoevaluados. Si no se tienen en cuenta estos factores, «nunca se puede estar seguro de si el alto grado de satisfacción del cliente está relacionado más con factores como la carencia de conocimientos o unas expectativas limitadas, que con la ayuda real que supone el servicio social que se presta» (Rees y Wallace, 1982, pág. 77).

La conclusión de todo esto no es que deba abandonarse el juicio sobre la satisfacción del consumidor. Los enfoques de la indagación narrativa para conocer la satisfacción superan algunos de estos problemas, porque permiten que los participantes en la evaluación eviten la tendencia a prescindir del contexto en los cálculos del estudio. Pero las mediciones de la satisfacción no son una panacea. Son difíciles de interpretar y de utilizar, y pueden resultar políticamente empobrecedoras.

Las indagaciones sobre la satisfacción se suelen realizar a pequeña escala y desde dentro, y son especialmente proclives a enfrentarse a los problemas de que hemos hablado. Algunos autores han señalado cuál es la buena práctica evaluativa para valorar la satisfacción (Pryce-Jones, 1993; Shaw, 1996a). Por ejemplo, los evaluadores no deben apoyarse en los índices globales de satisfacción. Ésta se ha de entender siempre a la luz de las *aspiraciones* y las *expectativas* contextualizadas. Puede ser útil distinguir entre expectativas preferidas (o ideales) y expectativas prácticas (o realistas). Una respuesta general que diga «satisfecho» puede significar cualquier cosa de las siguientes:

- He evaluado el servicio y estoy contento con él.
- No creo que tenga capacidad para evaluar, pero confío en el personal.

- El servicio era horrible, pero no me gusta criticar; al fin y al cabo, hacen todo lo que pueden.

Una última advertencia. La satisfacción del consumidor, sin más pruebas, normalmente no se debe considerar un indicativo de éxito. Resulta tentador, e injusto, sacar esta conclusión simplemente porque hemos conseguido que la gente se sienta contenta con su experiencia. Fisher llega a decir que:

Los índices de satisfacción [...] no se pueden tomar como signo de que el servicio consigue alcanzar sus objetivos o los de los clientes. Estas mediciones se refieren principalmente a la calidad del encuentro entre el trabajador social y el cliente, no a su resultado (1983, pág. 42).

La evaluación de potenciación y del profesional

Abriamos este capítulo con un ejemplo de la interacción entre el método y el propósito de la evaluación participativa y de potenciación. Más que volver a este tema, quiero intentar fijarme con mayor atención en lo que dicen los profesionales sobre su evaluación. En los párrafos siguientes, Searight expone la observación y el análisis que hizo de su práctica como terapeuta. Miller describe un proyecto en el que trabajó con otras personas, destinado a extender el concepto de profesor como investigador a formas recíprocas e interactivas. Searight tiene sus reservas acerca de la utilidad más amplia del modelo que ella siguió, mientras que Miller cree que en el trabajo que ellas y sus colegas realizaron existen posibilidades transformadoras.

EJEMPLO 8.4. AUTOEVALUACIÓN DE LA TERAPIA FAMILIAR CON NIÑOS QUE SUFREN TRASTORNOS DE CONDUCTA

En su trabajo con once familias, Searight (1988) utilizó una forma de observación del participante, entrevistas informales, y una mezcla de registros en audio y de relatos. Sus intereses eran: qué relación existe entre los contextos y la autoevaluación; la importancia de la participación del terapeuta en el mundo del cliente; las diversas perspectivas de los profesionales, los coterapeutas y los clientes; y de qué forma los significados que asumen esas personas configuran y caracterizan la autoevaluación.

Partió del supuesto de que el investigador necesita participar en el mundo de significados del sujeto, y de que esto no implica necesariamente una participación en el entorno del cliente. Observó hechos directos de la terapia, incluidos algunos hechos observados en los hogares y en la comunidad. Observó todas las sesiones de terapia durante un período de dieciséis meses.

Su propia perspectiva cambió, aunque no consiguió interpretar de cierto modo estos cambios hasta que no realizó el pertinente análisis. En efecto, un tema esencial era que «el profesional-evaluador no se podía centrar en la recogida de datos y simultáneamente participar de forma completa en las evaluaciones de la práctica» (Searight, 1988, pág. 312). Las conclusiones de la autora van aún más allá. Realizó la mayor parte de los análisis después de abandonar el enclave en el que hizo las observaciones, y llegó a la conclusión de que «el examen del contexto, en especial el organizativo y de la entidad, parece particularmente difícil mientras se está en el enclave» (1988, pág. 314).

Sin embargo, la utilidad del modelo está en que obliga al profesional a dar una definición de la práctica que no sólo incluya las destrezas, sino también los contextos de esa práctica. Además de sus beneficios analíticos, este enfoque «ayuda al trabajador social a alejarse de una práctica excesivamente personalizadora» (1988, pág. 328) y a evitar que personalice la culpa cuando las cosas van mal. «En la evaluación de la práctica, los contextos son al menos tan importantes como las destrezas» (1988, pág. 330). Esto supone contrarrestar de algún modo un defecto reconocido de la autoevaluación: el de que muchas veces no sabe formular preguntas críticas ni situar los proyectos en marcos sociales más amplios.

El discurso de la potenciación defiende la idea de que las personas producen el cambio dentro de su entorno y en él, y de que ellas no son el objeto de las estrategias de producción del cambio. Miller y sus colaboradores registran un ejemplo convincente de este tipo de indagación, que lleva la investigación del profesional a un terreno diferente (Miller, 1990, 1992; Miller y Martens, 1990).

EJEMPLO 8.5. PROFESORES QUE COLABORAN EN LA POTENCIACIÓN

Miller trabajó con un grupo de sus antiguos alumnos, todos ellos dedicados a la enseñanza. El grupo trabajó durante al menos seis años. Hacían transcripciones de lo que se decía en sus reuniones, y ca-

da uno de los miembros convino en llevar un diario, con lo que «se desbarataban las narrativas totalizadoras que anulan la diferencia y la contradicción» (Miller y Martens, 1990, pág. 47).

Pronto surgieron recelos ante la expresión «investigación emancipadora», como un tipo de «discurso autoritario» (Miller, 1990, pág. 16) de otra persona. Hubo un esfuerzo constante por «no imponer el concepto de potenciación a los demás en nombre de la liberación» (1990, pág. 18; Miller, 1992). La colaboración de esas personas tuvo tres limitaciones. Primera, necesitaban determinación y apoyo mutuo para encontrar tiempo y lugar para reunirse y seguir los diarios. Segunda, pasaron temporadas trabajando con diversos «niveles de expectativas» (Miller, 1990, pág. 70). La narración constante de sus historias impidió que el grupo fuera presa de una única narrativa. Tercera, tenían que vivir con la incertidumbre. El proceso provocaba unas «discrepancias privadas pero desestabilizadoras en su rol de educadores» (1990, pág. 77) y un deseo de obtener la respuesta correcta. Éstos eran algunos de los temas recurrentes a los que se enfrentaban:

- El miedo a la posibilidad de que «no hubiera nada».
- Una constante reaparición de «conceptualizaciones del conocimiento determinada por las investigaciones de otra persona» (1990, pág. 71). La separación habitual entre investigación y práctica «confundía nuestros esfuerzos por concebir la investigación como una *praxis*» (1990, pág. 73).
- La incertidumbre, en palabras de uno de los miembros del grupo, de saber: «¿Lo estamos haciendo bien?».
- No poder hablar con nadie que no fuera del grupo y que compartiera sus supuestos de convertirse en creadores más que en transmisores de conocimientos objetivos. La escuela tradicional no fomenta ni acepta investigar o reflexionar sobre el trabajo. Lo curioso de la investigación de ese grupo era que sus miembros formaban parte de una institución que perpetuaba el modelo de racionalidad técnica de investigación y evaluación, un modelo que esas personas deseaban cuestionar.
- «La naturaleza vacilante y ondulante de nuestra voluntad de desafiar lo establecido» (1990, pág. 70).

El grupo hizo frente a «las tensiones entre nuestro deseo de comunidad y apoyo en estas exploraciones, y los peligros de una unidad colectiva que podía ocultar las diferencias que existían entre nosotros» (1990, pág. 150). Reflexionaban también sobre la importancia de los cambios «pequeños». Uno de los del grupo decía:

Me di cuenta de que, en última instancia, realmente no me importaba el producto final, la planificación de las clases. Lo que realmente quería era el proceso, el forcejeo con la confusión, la identificación de las discrepancias [...] Imagino que parecerá un paso pequeño, pero para mí es muy grande, ya que siempre me he servido de las recetas (1990, págs. 138-139).

El grupo se mostraba escéptico ante la idea del profesor como investigador: «Nos preguntamos si esta expresión, profesor como investigador, contribuye de alguna forma sutil a esas distinciones teóricas y prácticas que siguen confundiendo la idea de que la propia enseñanza es una "forma silenciosa de investigación"» (1990, pág. 150). Les preocupaba también que la investigación del profesor cuestiona muy poco, o nada, la naturaleza a menudo jerárquica de las investigaciones y las interpretaciones que desde la universidad se hacen del trabajo de aula de los profesores (Miller y Martens, 1990). Más tarde, Miller reflexiona sobre los profesores y los administradores de aula, y dice:

Los principales modos de análisis y de indagación autorreflexiva en la mayoría de ellos no se encuentran depositados en los discursos académicos, la metodología cualitativa, ni en los libros, sino en la enseñanza que imparten todos los días, en la función de administración y de asesoramiento que ejercen en sus respectivos centros (Miller, 1992, pág. 170).

Así pues, cuando el grupo hablaba del poder, «no se esforzaban directamente por tratar los temas del poder y de la autoridad dentro de la investigación cualitativa. Lo que pretendían era hablar de las situaciones de poder y de autoridad dentro de las escuelas» (1992, pág. 171).

Heron emplea el lenguaje de Nietzsche cuando distingue entre culturas apolíneas y culturas dionisiacas de la indagación cooperativa. La indagación apolínea sigue un ciclo racional de secuencias y pasos: planificar, actuar, observar y reflexionar, y volver a planificar. La indagación dionisiaca adopta «un enfoque imaginativo, expresivo, en espiral, difuso, improvisado y tácito para tratar la interacción de la interpretación y la acción» (Heron, 1996, pág. 46). Heron sitúa su trabajo en la cultura apolínea. El trabajo de Miller y sus colegas es un valioso incentivo para exponerse a las incertidumbres de la evaluación colaborativa dionisiaca. El éxito de ese grupo en su compromiso compartido, bien que en un contexto de muchas incertidumbres, es impresionante.

La ética del trabajo de campo

Los dos ejemplos anteriores presumen el carácter omnipresente de los temas éticos en el trabajo de campo. Lofland y Lofland (1995) hacen una relación de las principales cuestiones éticas que previsiblemente se van a plantear en la investigación cualitativa:

- ¿Es ético observar que existe una necesidad grave de ayuda y no reaccionar directamente?
- ¿Es ético el hecho de adoptar una postura calculada ante otros seres humanos?
- ¿Es ético tomar parte o evitar tomar parte en una situación de división?
- ¿Es ético «recompensar» a las personas por permitir que se acceda a su vida y a su mente?

Cada uno de estos problemas encuentra una determinada expresión en la relación que se establece con la evaluación. Los estudios feministas, más que cualesquiera otros, han conseguido situar en primer plano estos temas (Acker y otros, 1983; Ellsworth, 1989; Stacey, 1988). Las investigaciones de Acker, Barry y Esseveld versaban sobre la relación entre los cambios en la situación del mercado de trabajo de las mujeres y los cambios de la conciencia. Probaron tres estrategias para reducir el poder desigual y reconocer la subjetividad de los participantes, aunque admiten que no resolvieron este problema:

1. Fomentar que el entrevistado tomara la iniciativa a la hora de decidir de qué hablar. No siempre funcionó. Las mujeres solían preferir que se les hicieran preguntas. Sin embargo, el sistema funcionaba mejor con las mujeres a las que las autoras realizaban más de una entrevista.
2. Establecer una reciprocidad, y para ello las autoras se ofrecían a contar algo propio a las mujeres. Se aceptaba y solía invitar a que éstas hicieran lo mismo, e «hicimos amistad con muchas de las mujeres» (1983, pág. 428):

Sin embargo, observábamos una tensión normalmente no articulada entre la amistad y el objetivo de la investigación. La meta de la investigación siempre es recoger información, por lo tanto [...] intentar crear una relación más igual paradójicamente puede convertirse en una explotación y un abuso (1983, pág. 428).

3. Mostrar el material escrito de las autoras a las mujeres. Sin embargo, no lo hacían con todas ellas. Admiten que lo comparían sobre todo con «aquellas que consideraban que conscientemente intentaban cambiar». Por consiguiente, eran las mujeres que «más compartían nuestra visión del mundo», y «tenemos que admitir que teníamos cierto reparo en compartir nuestras interpretaciones con aquellas que esperábamos que se iban a molestar por su contenido» (1983, pág. 428). Esto plantea el tema de si el evaluador debe compartir las interpretaciones con aquellos cuyas explicaciones son radicalmente distintas. Es

la tensión entre el objetivo de reducir las diferencias de poder entre investigador e investigado, y las dificultades de hacerlo cuando no hay acuerdo sobre el significado de las experiencias (1983, pág. 429).

Los temas éticos se han de contextualizar, con independencia de cuál sea nuestra postura sobre la relatividad y el relativismo. Un último ejemplo ilustra la forma que puede adoptar todo esto. Está sacado de una mesa redonda sobre ética, que tuvo lugar en un simposio sobre indigencia y exclusión social que en 1996 organizó el Conseil National de L'Information Statistique, de París. En el ejemplo he conservado el formato de mesa redonda, sin desvelar la identidad de los participantes.

EJEMPLO 8.6. ÉTICA E INVESTIGACIÓN DE LA INDIGENCIA

Presidente: Dijo, en resumen, que los temas éticos en este campo están ligados a:

1. El propósito de cualquier indagación.
2. El equilibrio existente entre el valor de la indagación y el grado de intromisión.
3. ¿Quiénes son los beneficiarios? ¿El entrevistado es un beneficiario?
4. ¿Cómo se obtiene la información?
5. ¿Cómo se va a utilizar? Las cuestiones de intimidad y la posible mala interpretación de las conclusiones.

El **participante 1** sostenía que los temas éticos no son *más importantes* en la investigación sobre la indigencia, pero necesitan una *formulación más meticulosa*. Pensaba que entre ellos están:

- El derecho a negarse en *el momento del contacto*. Pero este derecho no se puede ejercer en nombre de otra persona. Estaba en contra del derecho de cualquier grupo a tomar en nombre de otros la decisión colectiva de no participar.
- *La entrevista*. Es necesario disponer de un espacio privado que garantice la intimidad. Esto puede suponer una dificultad especial en los estudios sobre indigentes.
- *El contenido de la entrevista*. Este tipo de entrevistas suelen poner el énfasis en lo negativo, en las deficiencias del entrevistado.
- *La conclusión de la entrevista*. Es difícil ofrecer alguna compensación, ya que no existe una dirección donde remitirla, pero es un tema fundamental.

El **participante 2** habló de un posible pago a los entrevistados. Es algo muy raro en las investigaciones que se realizan en Francia. Por un lado, plantea el problema del paternalismo. Pero el hecho de que uno acepte que le entrevisten puede significar que pierda el turno en la cola de la comida o el de pedir limosna. La solución fue dar una tarjeta telefónica al finalizar la entrevista, pero sin decir de antemano que se iba a dar. Con la tarjeta se puede hablar con familiares y amigos, pero no se puede utilizar directamente de forma que uno salga perjudicado.

Al **participante 3** le interesaba la ética de los estudios longitudinales que se realizan en Estados Unidos. Habló de un estudio en el que se hicieron 1.500 entrevistas, de 500 de las cuales se hizo un seguimiento en intervalos de dos meses, durante dieciséis meses. Las cuestiones éticas que planteó fueron:

- *Pagos*. En Estados Unidos es habitual que en las investigaciones se paguen las entrevistas largas. Reconocía que dar dinero podía tener sus riesgos, pero defendía una idea no paternalista de la remuneración.
- *Intimidación*. Existe el problema de saber cómo se va a utilizar la información. Por ejemplo, para un indigente podría ser perjudicial que se pidiera información a un tercero (quizás un funcionario) relativa a la situación de indigencia de esa persona. Podría poner en peligro la posibilidad de que el organismo que diera la información siguiera ofreciendo el «servicio». Como solución parcial, daban a la investigación un título que fuera inocuo. En vez de llamarla «Estudio sobre la indigencia», lo llamaban «Estudio sobre la disponibilidad de vivienda».
- Las implicaciones y las *obligaciones* derivadas de mantener una prolongada relación con las personas.

El participante 4, un trabajador social, distinguía entre quienes están «condenados a dar» (los asistentes sociales) y los que están «condenados a recibir». «La vida les ha puesto los cuernos a los indigentes.»

Lo que importa es lo que ocurre *después* de la investigación. ¿Cómo podemos corresponder y devolverles los conocimientos que nos proporcionaron? El tema ético clave es aprender a recibir. El invitado reconocía que no disponía de una auténtica respuesta a este problema de corresponder a lo recibido. Había participado en la organización de una exposición de arte para gente sin hogar y los beneficios que se obtuvieron fueron destinados al uso colectivo. Destacaba que también correspondemos cuando nos guardamos la información. Escuchar, como responder, forma parte de esa correspondencia. A veces no se puede localizar a la persona o es posible que haya muerto, pero debemos incorporar a nuestra conducta una actitud «atenta».

El participante 1 decía que la entrevista reporta realmente un beneficio. Puede capacitar a las personas para que no vean la vida como un destino inamovible (por ejemplo, aprendiendo a leer entre líneas). Esto debería formar parte de nuestras preocupaciones éticas. Por último, no debemos olvidar los beneficios colectivos de la investigación. Los juicios sobre beneficios no se deben limitar al beneficiario inmediato.

Los temas fundamentales de este capítulo han sido la articulación del propósito y el método de la evaluación, la contextualización local de las decisiones sobre métodos concretos y las cuestiones que su uso plantea. Estos mismos temas se aplican al análisis de las pruebas de la evaluación, un asunto al que dedicamos el último capítulo.

CAPÍTULO 9

EL ANÁLISIS: ESTABLECER LAS AFIRMACIONES DE LA EVALUACIÓN

El análisis de los datos no es un elemento diferenciado del proceso de evaluación que se pueda separar perfectamente. En efecto, hemos recorrido casi un círculo completo. En el capítulo introductorio reflexionábamos sobre la dimensión *evaluativa* de este libro, y también sobre en qué sentido trata de la evaluación *cualitativa*. En este último capítulo hablamos de cómo dar un fundamento a lo que se afirma en la evaluación cualitativa. Puede resultar casi tedioso que diga una vez más que no creo que exista un acuerdo global del que nos sirvamos. La evaluación se distingue por los fines que la identifican, más que por el uso de una metodología particular. Al igual que en el diseño y el trabajo de campo de la evaluación, las estrategias de análisis estarán configuradas por esos propósitos particulares.

El conjunto de estrategias de las que se sirve el análisis son muy similares a las que existen para la investigación cualitativa. Por esta razón, en vez de hurgar en los estantes en busca de libros que traten del análisis de la evaluación, es mejor aconsejar a los futuros evaluadores que empiecen con los textos más recientes sobre el análisis de la investigación cualitativa (Coffey y Atkinson, 1996; Lofland y Lofland, 1995; Miles y Huberman, 1994; Silverman, 1993) y que después filtren estas lecturas según sean las necesidades de la investigación que realicen. En la primera parte de este capítulo, reflexionaremos brevemente sobre el análisis de las entrevistas cualitativas, no porque pensemos que éstas sean un modelo para *todos* los análisis, sino como ilustración de algunos temas y procesos analíticos y para abordarlos.

La parte más importante del capítulo trata de los temas que plantea el análisis de la evaluación cualitativa. Recogemos el hilo del capítulo 6 y señalamos de qué forma la evaluación *del profesional* y la *participativa* configuran el sistema de análisis. El análisis participativo plantea importantes cuestiones de *ética del análisis*, y nos ocupamos de éstas y de otras consideraciones que se suscitan al analizar

una evaluación y al informar sobre ella. Esto facilita el contexto para abordar la cuestión fundamental del capítulo: cómo se pueden fundamentar y respaldar las afirmaciones de la evaluación. Volvemos a la tesis del capítulo 7 sobre la relevancia de la evaluación cualitativa para ocuparse de las espinosas cuestiones de las *causas*. De ahí pasamos a considerar algunas afirmaciones sobre la forma de hacer más verosímil el análisis cualitativo. Entre otras posibilidades están el uso de la *validación de los miembros*, la *triangulación*, y los métodos de *síntesis* y de *metaanálisis*. Por importantes y difíciles que puedan resultar estas cuestiones, probablemente lo sean más los temas de la *generalización* y de la *validez externa* de la evaluación. Señalamos y revisamos las opciones y los enfoques de mayor aceptación. Terminamos con un último repaso de los temas relacionados con la *autoría*, la *información* y el *uso* de la evaluación, de los que nos ocupábamos en un sentido más general en el capítulo 4.

LAS AFIRMACIONES DE LA EVALUACIÓN

Lo esencial de la evaluación es «el intento de establecer y mantener la posibilidad de defender ante los clientes y los interesados las afirmaciones normativas y empíricas sobre ciertos fenómenos» (Fournier y Smith, 1993, pág. 316). Sin embargo, la lógica deductiva formal, basada en unas pruebas inequívocas, no puede servir de base de tales afirmaciones.

Todo el mundo está de acuerdo en que la información de algún modo conforma las decisiones, pero la relación no es directa, ni sencilla. A menudo cuanto más importante es la decisión, más oscura parece ser la relación (House, 1980, pág. 68).

House llega a decir que «si se someten a un examen riguroso, las evaluaciones siempre parecen equívocas» (1980, pág. 72). Ya hemos visto antes la idea de este autor de que las evaluaciones no pueden ser más que actos de persuasión: «La evaluación, más que convencer, persuade; más que demostrar, argumenta; más que cierta, es creíble; más que convincente, se acepta de forma variable» (1980, pág. 73). Muchas veces los evaluadores han restado importancia al papel que desempeña el enjuiciamiento y, por consiguiente, a la argumentación. El resultado ha sido un análisis excesivamente técnico y orientado a los métodos, una confianza exagerada en la cuantifica-

ción y una tendencia a representar los conocimientos de forma impersonal y objetivizada. A quienes no aceptan las conclusiones «convincientes» que se sacan de la evaluación se les tacha de irracionales. Si los resultados son inequívocos, entonces quienes no saben aceptarlos están «equivocados».

Ha habido varios intentos de enfocar de distinta forma el pensamiento a través del proceso de razonamiento que supone la construcción de unos argumentos evaluativos justificados (Fournier, 1995; Fournier y Smith, 1993; Scriven, 1997). En estos enfoques lo habitual es destacar la compleja conexión que existe entre las pruebas y las conclusiones, y parten de las diferencias entre el razonamiento formal y el informal. Si el razonamiento formal presume un perfecto ajuste entre las premisas y las conclusiones de una argumentación, la lógica informal «trata del lenguaje corriente y cotidiano, donde las reglas de la inferencia son menos precisas y más contextuales» (Fournier y Smith, 1993, pág. 317). *The Uses of Argument* (1958), de Toulmin, ha sido quizás el texto de mayor influencia. Este autor, al explicar la conexión entre las afirmaciones y las pruebas, otorga un papel fundamental a las «justificaciones». Una justificación proporciona la base para explicar cómo pasamos de la prueba a la afirmación y, por lo tanto, para legitimar la inferencia que realizamos. En apartados posteriores de este capítulo nos ocuparemos de algunas justificaciones características de las que se sirven los evaluadores cualitativos.

Este planteamiento tiene varias ventajas. Por ejemplo, en vez de empezar en argumentos teóricos sobre cómo deberíamos razonar, nuestra atención se dirige a averiguar cómo razonan realmente en el campo quienes practican la evaluación. Los estudios empíricos de estos procesos nos permitirían comprender si las inferencias de la evaluación difieren, y cómo difieren, en función de los diferentes propósitos, enclaves, metodologías, etc. Por ejemplo, los diversos autores cuya obra presentábamos en el capítulo 2 emplean unas justificaciones distintas. Las argumentaciones de Fournier, Scriven y Smith se ajustan mejor a lo que hemos denominado una postura multipartidista. Eraut utiliza la argumentación exactamente de esta forma (Eraut, 1984, pág. 33), una forma menos compatible con las ideas de ilustración o explicación del uso de la evaluación, o con las posturas de la defensa comprometida. Sin embargo, el carácter inductivista y de orientación empírica del enfoque abre un camino prometedor. Plantea a los evaluadores una serie de preguntas críticas con las que examinar su práctica:

¿Qué tipo de afirmación, o afirmaciones, intento hacer en esta evaluación? ¿En qué fuente de autoridad debe asentarse la justificación? ¿Qué fuerza tiene la fuente de autoridad para respaldar las inferencias realizadas a partir de las pruebas? ¿Qué condiciones socavarían la justificación? ¿Hasta qué punto es, o será, convincente la conclusión resultante? (Fournier, 1995, pág. 28).

LAS JUSTIFICACIONES CUALITATIVAS

Hace algún tiempo, Cronbach insistía en que proceder sobre la base de que la evaluación parte de un acuerdo sobre los objetivos es algo «racionalista hasta el punto de resultar irreal» (Cronbach y otros, 1980, pág. 129). Todos los programas sociales, decía, tienen unos objetivos amplios y vagos, incluso los programas que supuestamente apuntan a un objetivo concreto. Esto no significa ser escapistas, como dirían los racionalistas, sino que refleja la naturaleza de los programas, que funcionan en un ambiente de acuerdo político: «La primera regla del buen proceso político es: "No forzar la concreción de unas metas o unos fines"» (1980, pág. 130).

La lección se ha asimilado. Silverman, en su estudio de la orientación sobre el sida, sigue encontrando razones justificadas para lamentar que lo que él denomina la Ortodoxia Explicativa de los estudios de orientación lleve a centrarse o en las causas o en las consecuencias de la orientación. En este planteamiento «hay tanta prisa por llegar a una explicación que es imposible hacer preguntas serias sobre lo que se explica» (Silverman, 1997a, pág. 24), de tal forma que el fenómeno de la orientación «se escapa». Las formulaciones tradicionales de las aportaciones y los resultados causales se deben posponer al menos hasta que se comprenda el «cómo». Por razones evaluativas bien fundadas querremos plantear las preguntas explicativas del «por qué»: «No hay razón para no hacerlo, dado que primero hemos descrito con precisión cómo se produce localmente el fenómeno en cuestión» (1997a, pág. 35). Un enfoque cualitativo del análisis de la evaluación partirá de la premisa de que «los fines de los objetivos deben surgir del proceso, y no al revés» (Sherman, 1990, pág. 152).

Pero pese a todas las virtudes de la investigación de evaluación cualitativa, los responsables de las políticas y de las decisiones que se hayan formado en la tradición dominante la criticarán de forma repetida, alegando que:

- Es una evaluación a demasiado pequeña escala.
- No es representativa y, por lo tanto, no se puede generalizar.
- Al buscar la validez en el nivel del significado, se queda corta en lo que se refiere a la adecuación causal (Finch, 1986).

No hay razón para que los evaluadores acepten estas críticas sin protestar. Finch se muestra moderada y poco amiga de la polémica cuando concluye que la investigación cualitativa de políticas tiene las ventajas de:

La flexibilidad del propio proceso de investigación; unos estudios que sitúan la vida social en su contexto natural; una preocupación por el proceso, además de por los resultados; unas explicaciones que son adecuadas en el ámbito del significado, pero en las que se consideran también las cuestiones de la adecuación causal, aunque no se puedan resolver por completo (1986, págs. 161-162).

Todos los métodos de trabajo de campo que hemos esbozado en los capítulos anteriores pueden favorecer este tipo de fines del análisis evaluativo, unos fines que van desde el análisis de los microprocesos mediante sistemas como el análisis de conversaciones (Silverman, 1997a) o el estudio del proceso de cambio (Reid, 1990), hasta el análisis de alcance más amplio que se asocia con el estudio de casos (Burgess y otros, 1994; Stake, 1995; Yin, 1994). Entre lo uno y lo otro se sitúan, por ejemplo, los nuevos trabajos sobre el análisis de los grupos centrados (Frankland y Bloor, 1999). Vamos a repasar brevemente, con ejemplos, algunos de los temas y los procesos analíticos que implica el análisis de las entrevistas cualitativas.

EL ANÁLISIS DE LAS ENTREVISTAS DE LA EVALUACIÓN CUALITATIVA

Existe una rica literatura sobre el análisis de la entrevista cualitativa (Crabtree y Miller, 1992a; Holstein y Gubrium, 1995; McCracken, 1988) y una diversidad de estilos analíticos. Los estilos analíticos se ubican en algún punto de un continuo que va de lo deductivo a lo inductivo. El *análisis de contenidos* básico adopta un enfoque sistemático y deductivo y aporta un sentido teórico previo claro de los problemas analíticos a un análisis de los contenidos sustantivos del texto. La obra de Reid (1990) sobre la investigación del proceso de cambio, como un medio para la evaluación del desarrollo, es un ejem-

plo de ese enfoque. Otro planteamiento es desarrollar un sistema de codificación previo basado en investigaciones anteriores y en las ideas y la experiencia del evaluador, con el fin de abordar los contenidos latentes del texto de una forma relativamente sistemática y deductiva. Ilustramos este enfoque cuando pasamos a hablar del análisis causal en la evaluación. En tercer lugar, está lo que Crabtree y Miller describen como el «estilo editor» (1992b, pág. 211), que adopta un planteamiento inductivo y comparativo en el establecimiento de los temas y las tesis analíticas. Por último, el enfoque más inductivo del análisis implica que el evaluador llega a sumergirse en el texto con el propósito de materializar su interpretación esencial, de una forma «intuitiva y muy dirigida por los datos» (1992b, pág. 212).

Ninguno de estos enfoques se puede rechazar de plano en la evaluación. Incluso el sistema exclusivamente inductivo puede ser la opción preferida en los proyectos de evaluación etnográfica y de construcción de conocimientos. Una breve descripción del tercer enfoque puede servir para ilustrarlo.

El evaluador analiza:

- La impresión de lo que la literatura y los propios trabajos anteriores de uno mismo dicen que debería haber en los datos.
- La impresión de que este tema se constituye en la propia experiencia de uno mismo.
- La «impresión de lo que tuvo lugar en la entrevista» (McCracken, 1988, pág. 42).

El analista ha de empezar por identificar las declaraciones clave del texto de una entrevista y ha de estar preparado para «vislumbrar y reconstruir sistemáticamente una idea del mundo que no guarda relación» con su propia idea ni con la de la literatura. Esto supone cultivar una lectura «afectada» del texto: «Llegar al texto con cierto falso asombro, negándose a aportar los supuestos y las interpretaciones» (McCracken, 1988, pág. 44). Esto implicará observar cualquier asociación que el texto evoque en el lector, incluidas aquellas ocasiones en que el evaluador reconoce lo que se dice, no porque la declaración se haya decodificado con éxito, sino porque «ha tenido lugar un repentino acto de reconocimiento» (1988, pág. 44).

El entrevistador evaluativo pretende identificar los «vínculos narrativos». «La entrevista activa tiene dos objetivos clave: recoger información sobre *lo que* trata el proyecto y explicar *cómo* se construye narrativamente el conocimiento referente a ese tema» (Holstein y

Gubrium, 1995, pág. 56). Holstein y Gubrium insisten en los aspectos colaborativos de la construcción de significado dentro de la entrevista. Por ejemplo, describen la codificación como una actividad «indígena» que tiene lugar y se desarrolla como una parte integral del proceso de la entrevista, y no sólo antes o después.

EJEMPLO 9.1. LA CODIFICACIÓN INDÍGENA EMERGENTE

En un ejemplo de lo que Fournier (1995) describiría como evaluación «de establecimiento de justificaciones», una funcionaria encargada de la vigilancia de las personas que están en libertad condicional expone un caso de su trabajo que no fue bien. Un hombre estaba en libertad condicional después de haber amenazado de muerte a su esposa poniéndole un cuchillo en el cuello. Una vez puesto en libertad condicional parecía que iba «increíblemente bien». La funcionaria encargada de su control estaba «dispuesta a remitir la orden al tribunal con la solicitud de que se revocara, porque pensaba que habíamos progresado mucho [...] Luego, claro, su mujer me llamó por teléfono y [...] dijo que él no había dejado de beber y de acosarla».

Una pregunta del entrevistador codifica la realidad al pedirle a la funcionaria que piense desde la perspectiva de determinadas opciones evaluativas. En respuesta a una pregunta del entrevistador, la funcionaria se refiere a diversos «códigos» analíticos para comprender su trabajo. Algunas de las expresiones clave se ponen en cursiva en el fragmento siguiente.

ENTREVISTADOR: ¿Qué había en ese caso que le hacía pensar que no estaba haciendo bien las cosas?

FUNCIONARIA: (*Suspiro*) Era evidente que me había creído *sin más* lo que el hombre decía. Visto ahora, no sé si me debería haber esforzado más en hacer un seguimiento con su ex mujer. Me refiero a que, *en ese momento*, estaba (*suspiro*) [...] me parecía que bastaba con informar a la mujer, dado que ya no estaban casados, él no vivía con ella y ésta tenía *derecho a su intimidad* y a que no se la involucrara. Había también una cuestión de *confidencialidad* sobre lo que el hombre me contaba en las entrevistas, algo que obviamente no podía compartir con su ex mujer sin su permiso. O sea que me encontraba... quiero decir que básicamente me sentía como una auténtica idiota, en lo que a la efectividad se refería, no sólo completamente ineficiente, sino que posiblemente *empeoraba una mala situación* porque *actuaba en connivencia* con ese hombre creyendo que todo iba bien de verdad, y *desperdiciaba un montón de tiempo* en unas sesiones supuestamente de orientación (*risas*).

Además de este tipo de códigos indígenas, el entrevistador también pretende fomentar la visibilidad de los vínculos narrativos. Holstein y Gubrium llegan a decir que «el entrevistador desafía interpretativamente al entrevistado para comprender la experiencia relativa a diversas posibilidades subjetivas» (1995, pág. 59). La secuencia de pregunta y respuesta que sigue inmediatamente después del intercambio del ejemplo 9.1 ilustra la forma que esto puede adoptar.

EJEMPLO 9.2. FOMENTAR LOS VÍNCULOS NARRATIVOS VISIBLES

ENTREVISTADOR: ¿Es que parte de la dificultad es la propia dificultad de poder interpretar las pruebas?, porque lo que usted decía que había parecían ser montones de indicios que razonablemente se podía esperar que usted interpretara que significaban que las cosas iban bien.

FUNCIONARIA: Sí.

ENTREVISTADOR: ¿Y de repente unas pocas pruebas ponen en peligro todas las demás?

FUNCIONARIA: Eso es. Quiero decir que es una prueba concreta, la llamada de su ex mujer, la que lleva a una interpretación completamente distinta de todo lo demás que había ocurrido.

ENTREVISTADOR: ¿Tiene que ver esto en general con la interpretación del trabajo, con el problema de interpretar las pruebas?

FUNCIONARIA: Es difícil evaluar las actitudes de las personas (*risas*) porque dependes mucho de la información que quieran darte y de lo que te cuenten. Ese tipo, por razones que él sabrá, decidió confundir deliberadamente. Asistía a las citas con escrupulosidad, era puntual (*risas*). Se esforzaba mucho por dar una determinada imagen. Por las pruebas de que yo disponía hice una evaluación que resultó ser completamente falsa, debido a las pruebas a las que no tuve acceso, es decir, por no saber qué ocurría realmente cuando visitaba a su ex mujer.

Evidentemente, la labor analítica supone mucho más de lo que aquí hemos expuesto. Los pasos que van de las pruebas a las afirmaciones de la evaluación son complejos, como ya hemos señalado. A través de la aparición progresiva de los temas analíticos, el objetivo es comparar y contrastar las categorías emergentes y transformarlas en categorías más amplias o tesis. Crabtree y Miller describen el resultado de un proceso similar en su análisis de un estudio en el que se comparaba la percepción que del dolor y de la gestión de ese dolor tenían los médicos y los pacientes (1992b, págs. 213-218). La tesis analítica de los autores es la siguiente:

1. Tanto el médico como el paciente tienen una visión personal del dolor, que deriva de su experiencia personal y que apoya sus identidades personales.
2. Ambos tienen también una visión profesional del dolor, pero estas visiones proceden de fuentes distintas.
3. El médico aprende de su profesión y emplea esa visión para respaldar su identidad profesional.
4. El paciente aprende del contacto con los profesionales y emplea esta visión para respaldar su identidad personal.

En resumen, «el médico tiene unas identidades potencialmente en conflicto, mientras que las del paciente son armoniosas» (1992b, pág. 218).

Digamos una vez más que estos fragmentos de análisis de entrevista no se deben interpretar ni como exclusivamente evaluativos ni como representativos de toda la empresa analítica. Se proponen como ejemplos relevantes de análisis cualitativo, que demuestran el interés de la evaluación cualitativa por la interacción de los procesos sustanciales y constructivos.

Cambiamos ahora de objetivo y pasamos a considerar las cuestiones relacionadas que se plantean al establecer y fundamentar las afirmaciones evaluativas. Entre estas cuestiones están los temas del análisis causal, la validación de los miembros, la triangulación, la generalización y la redacción y el uso de la evaluación. Pero antes de ocuparnos de todos ellos, nos referiremos al análisis participativo y del profesional, y a las cuestiones éticas que el análisis plantea.

EL ANÁLISIS PARTICIPATIVO Y DEL PROFESIONAL

La evaluación participativa y del profesional tiene su propia incidencia en la construcción de significado a partir de la indagación. La práctica y la indagación *no* son lo mismo, y ya nos quejábamos antes de lo que a menudo se dice sin razón al respecto. Pero esto no significa en absoluto que sea inútil compararlas. Lang habla de las «diferencias en las estrategias de procesamiento de datos en la práctica del trabajo social y en la investigación cualitativa y [...] cómo se podrían integrar ambas» (Lang, 1994, pág. 265). Este interés por «las correspondencias y las discrepancias entre las actividades del investigador cualitativo y las del profesional de la asistencia social» (1994, pág. 266) se basa en el interés que la autora tiene por «posibilitar que tanto el

saber como el hacer deriven de los mismos datos» (1994, pág. 274). Lang ejemplifica de qué forma la evaluación del profesional tiene las mismas probabilidades de ocuparse de una práctica mejor como de una mejor evaluación. Por ejemplo, aboga por una mejor disposición de los trabajadores sociales para mantenerse en el patrón inductivo de la indagación. En una paradoja verosímil sostiene que «la teoría existente debe tener un estatus más provisional, un objetivo menos central [...] para abrir la posibilidad de construir una teoría» (1994, pág. 276).

No me convencen los argumentos con que se defiende que la investigación es una cosa y la práctica otra, y que se deben mantener separadas (Hammersley, 1992). No son lo mismo, pero tampoco se pueden desenredar perfectamente. El análisis evaluativo que aborda la variedad de intereses de la evaluación del profesional que estudiábamos en el capítulo 6, estará configurado por la tensión y la interacción creativa de diversos propósitos.

Algo parecido se puede decir de los planteamientos participativos del análisis. El análisis suele pasar desapercibido en la investigación y la evaluación participativas. Muchas veces se cuestiona la primacía del método y la metodología cualitativa se mira con recelo. Tripp, por ejemplo, sostiene, a partir de una interpretación participativa de la entrevista, que «la estrategia de investigación ideal debería desarrollar la idea y a la vez facilitar información para evaluar su importancia para el entrevistado». Piensa el autor que esto no siempre debería ser una indagación educada, sino que incluirá «el debate de confrontación además de la discusión» (Tripp, 1983, pág. 34). «El objetivo es producir una explicación acordada de las ideas de los participantes, más que una exposición acordada de la discusión» (1983, pág. 35). Es una imagen radicalmente distinta tanto del estatus de los participantes en la entrevista como de los datos resultantes: «Los “datos” no son necesariamente las palabras concretas que se dicen durante la entrevista y que luego se transcriben, sino un documento que es el resultado negociado de las ideas de dos personas en el momento de la redacción» (1983, pág. 37).

La idea de Tripp de que los datos son «aquello que el entrevistado acepta de lo que se ha escrito sobre lo que se dijo» (1983, pág. 35) no puede «protegernos de nuestras propias pasiones» (Lather, 1986a, pág. 77). Heron tiene su propia versión del propósito de la evaluación. No limita los resultados de la indagación cooperativa a lo que él describe como «informes proposicionales», sino que incluye también:

- Las transformaciones del ser personal a través de la participación en el objetivo y el proceso de la indagación.
- Las exposiciones de ideas acerca del objetivo de la indagación a través de la danza, la representación y otros modos de expresión.
- Las destrezas prácticas que son 1) destrezas relacionadas con una acción transformadora en el ámbito de la indagación, y 2) destrezas relacionadas con diversos tipos de conocimiento participativo y de colaboración empleados en el proceso de la indagación (Heron, 1996, pág. 104).

Una vez que se acepta que el propósito de la evaluación incluye «actuar, y no simplemente afirmar» (Lather, 1991, pág. 123), entonces el análisis formará parte del empeño del evaluador por «maximizar el proceso de la investigación como un encuentro favorecedor del cambio y recíprocamente educativo» (1991, pág. 72), en el que el proceso es más importante que el producto.

Muchas de las obras feministas en este campo reconocen que el potencial emancipador del análisis participativo sólo se puede alcanzar en parte. Acker y sus colegas reflexionan en su estudio sobre la relación que se establece entre los intentos de las mujeres para integrarse en el mercado de trabajo y los cambios operados en la conciencia. En el capítulo anterior nos referíamos a las dificultades que las autoras encontraban en la relación de la investigación. Se enfrentaron a los consiguientes problemas en el análisis de su trabajo extensivo: «Nos encontramos con que teníamos que asumir el papel de personas con poder para definir» (Acker y otros, 1983, pág. 429). Por ejemplo, cuando pedían a las mujeres que leyeran su manuscrito, lo habitual era que éstas quisieran «más de nuestro propio análisis sociológico. Querían que nosotras, las investigadoras, les interpretáramos su experiencia» (1983, pág. 429). Llegaron a aceptar que:

En última instancia, el investigador debe objetivizar la experiencia del investigado, ha de traducir esta experiencia a términos más abstractos y generales, si se quiere conseguir un análisis que vincule al individuo con los procesos que se producen fuera de su mundo social inmediato (1983, pág. 425).

La tensión no se resuelve con la renuncia que Tripp hace de los datos en favor de los procesos de la negociación: «Si queríamos conseguir el objetivo de la emancipación para las personas que estudiábamos, teníamos que ir más allá de la representación fiel de su expe-

riencia, más allá de “dejar que hablaran por sí mismas”» (Acker y otros, 1983, pág. 429).

ÉTICA Y ANÁLISIS

La ingenuidad en lo que a la ética se refiere supone, en sí misma, una falta de ética. No debemos permitir que un adecuado énfasis en las diversas construcciones de la realidad y la distinción contextual local del programa, el proyecto o la práctica, justifiquen una indiferencia ética. La experiencia son «las historias que las personas viven. Las personas viven historias y, cuando las cuentan, las reafirman, las modifican y crean otras nuevas» (Clandinin y Connelly, 1994, pág. 415). Las cuestiones éticas se plantean por el hecho de que animamos a los usuarios de un servicio a que cuenten sus historias. Puede ser algo positivo y una forma de ayudar a alguien a valorar de nuevo un problema. Pero también conlleva peligros y destaca de nuevo que la evaluación se debe hacer con cuidado, y no como «una incursión en unas identidades extraviadas» (expresión de Dannie Abse, de su poema *Return to Cardiff*). Recuerdo que una mujer de una ciudad de un valle del sur de Gales se prodigaba al hablar de su experiencia de educar en casa a su hijo, que padecía unos problemas de aprendizaje graves y que, por entonces, tenía veinte años. Decía que era la primera vez que hablaba de su experiencia con otra persona.

Las cuestiones éticas más amplias se pueden ilustrar si se reflexiona sobre cómo las diferentes ideas de justicia afectan a las conclusiones que sacamos sobre si un determinado proyecto o programa es bueno o no. Una visión *utilitarista* de la justicia parte de la base de que las prácticas se juzgan correctas o equivocadas en función de si producen felicidad o satisfacción para la mayoría de las personas. La teoría de justicia social de Rawls se basa en la justicia como *imparcialidad*. Su famoso Principio de Diferencia establece que las desigualdades sociales y económicas son justas sólo si se traducen en una compensación de prestaciones para todos, en particular para los miembros menos favorecidos de la sociedad. Por consiguiente, se puede justificar que los salarios de los médicos o de los profesores se suban más de lo que suben los índices de prestaciones, alegando que de ello se pueden beneficiar las personas desfavorecidas. Quienes defienden una teoría *igualitaria* de la justicia insisten en que los desfavorecidos siempre han de obtener una prestación redistributiva, hasta que se consiga la igualdad. Consideremos el siguiente ejemplo de trabajo social.

Unos trabajadores sociales dirigen un grupo con el objetivo de reducir los delitos relacionados con los automóviles. Supongamos que el resultado es que el grupo tiene éxito para *todos* los miembros, pero *menos* en el caso de los que tienen un menor grado de alfabetización. Las conclusiones que sacamos dependerán de la idea implícita de justicia que tengamos (véase la tabla 9.1). Quizá merezca la pena analizar nuestra respuesta tácita a estos «resultados». Se nos ha socializado académicamente para responder con una inclinación metodológica. ¿Eran importantes las diferencias? ¿El programa se puso en práctica de forma adecuada? ¿El diseño ofrecía una buena base para esas inferencias causales implícitas? Son unas preguntas que pueden ser relevantes o no. Pero el aire metodológico con que tratamos las afirmaciones evaluativas ilustra que se ha preferido la eficacia antes que la justicia. Nuestra respuesta instintiva se convierte en parte del problema.

Pero esto puede parecer muy abstracto. Consideremos los dilemas éticos y políticos que plantea la forma de utilizar el material de la evaluación. Estas cuestiones a veces se agudizan más en la evaluación cualitativa que en el trabajo cuantitativo. Ello se debe en parte a la mayor proximidad y la consiguiente confianza que se pueden desarrollar entre el evaluador y el participante. En la investigación cuantitativa, una mayor distancia hará que esas cuestiones resulten menos angustiosas. También aumenta el peligro de traición, debido a que se suelen emplear muestras más pequeñas y a que se pone el énfasis en los detalles de la forma que las personas tienen de vivir su vida. Finch, en su investigación evaluativa de grupos lúdicos infantiles, habla de que tenía la «sensación de que existía la posibilidad de que traicionara a mis informantes como grupo, no como individuos» (1986, pág. 207). «Cuando la investigación cualitativa se centra en temas de política social, se plantea el particular dilema de que las conclusiones se empleen para *empeorar* de algún modo la situación de la población objeto del estudio» (Finch, 1985, pág. 117).

TABLA 9.1. *Justicia y resultados de la evaluación.*

Idea de justicia	¿Es un programa justo?
Utilitarista	Sí
Justicia como imparcialidad	Sí, si se beneficia el desfavorecido
Igualitaria	No, porque fomenta la desigualdad

EJEMPLO 9.3. GRUPOS LÚDICOS INFANTILES: ¿DE QUÉ LADO ESTAMOS?

Lo que a Finch le interesaba era dilucidar qué significaba la oferta de unos grupos lúdicos infantiles de autoayuda para las mujeres de clase trabajadora de zonas económicamente deprimidas. A lo largo de tres años, mediante la observación y las entrevistas semiestructuradas, la autora pudo documentar el carácter de los grupos lúdicos en esas áreas:

Descubrí situaciones en las que la práctica se apartaba completamente de los criterios sobre la educación y la atención a los niños que la mayoría de los responsables políticos y de los académicos tomaban como norma, y a veces eran muy peligrosas (1985, pág. 117).

Le preocupaba que la publicación de su trabajo reforzara más aún «esos supuestos profundamente anclados en nuestra cultura y nuestra política social, según los cuales las mujeres de clase trabajadora (en especial las pobres y urbanas) son unas madres inadecuadas» (1985, pág. 117). Quienes la habían aceptado durante tres años se iban a sentir traicionados.

Tuvo que abordar estos problemas. ¿Era culpable de haber impuesto a aquellos grupos lo que era la norma de la clase media? Y sin embargo, esta norma *era* a la que aspiraban las mujeres que dirigían los grupos. Era el modelo de las participantes, y no sólo el de la autora. Al final, ésta desarrolló un razonamiento que evitaba el modelo de «carencias» de la explicación y sostenía que considerar que las madres de clase trabajadora son incompetentes es un error y una ingenuidad. Reconoce que no está segura de haber resuelto por completo los problemas, y acepta que:

Sostener esto significa adoptar una postura francamente moral, muy alejada del modelo del científico social objetivo [...] Creo que la investigación cualitativa sobre cuestiones de política social llevará inevitablemente a unas posturas morales de ese tipo, y que nunca se puede limitar a ofrecer unos «hechos» (1985, págs. 119-120).

LAS CAUSAS Y LAS EXPLICACIONES

Para emplear el lenguaje cualitativo habitual, «los datos observacionales se usan para “generar hipótesis” o “describir procesos”. Los datos cuantitativos se usan para “analizar resultados” o “verificar hipótesis”» (Trend, 1979, pág. 83). En el capítulo 7 explicábamos con cierto detalle por qué estamos de acuerdo en que «esta división del

trabajo es rígida y restrictiva» (1979, pág. 83). Dicho esto, la división se refuerza con la frecuente ausencia de debate sobre la causa en los textos sobre metodología y análisis cualitativos (Coffey y Atkinson, 1996). La evaluación cualitativa no puede resolver los problemas de las conclusiones causales mejor que lo pueda hacer la evaluación cuantitativa, pero puede evaluar la causalidad «tal como realmente actúa en una determinada situación» (Miles y Huberman, 1994, pág. 10). Cronbach decía con un efecto similar que se necesitan conocimientos causales explicativos, «no conocimientos de las cosas que por norma general funcionan en una serie de condiciones diversas» (Shadish y otros, 1990, pág. 363).

La evaluación cualitativa comparte con la investigación cualitativa el reconocimiento de la ironía de las causas y las consecuencias sociales. Gran parte de la sociología de la desviación se basaba precisamente en este sentido de la ironía, con su exploración de los roles desviados que hacían el necesario «trabajo sucio». En términos evaluativos, podemos preguntar qué funciones cumple una determinada práctica y que no se cumplirían de no existir ésta. En términos menos funcionales podemos preguntar cuáles son los resultados típicos de este fenómeno en este enclave y a qué objetivos se sirve de este modo. Lofland y Lofland (1995) hacen la importante observación de que las respuestas causales se basan en general en unas concepciones *positivistas* de la naturaleza humana. La indagación cualitativa a menudo se ha apartado de las explicaciones causales, no porque la metodología sea deficiente en esta área, sino por un compromiso con una concepción *activista* de la naturaleza humana. Los Lofland sostienen que una concepción activista nos llevará a centrarnos en preguntas que consideren por igual las estructuras y las estrategias. Esto supondrá «descifrar y mostrar exactamente a qué tipo de situación se enfrentan los participantes» (Lofland y Lofland, 1995, pág. 146), y comprender las estrategias «fabricadas incesantemente» que las personas construyen para abordar la situación.

Pensemos, por ejemplo, en el trabajo de Silverman acerca de la orientación sobre el sida. Está en lo cierto el autor cuando concluye que «normalmente es innecesario permitir que nuestros temas de investigación se definan en términos de [...] las “causas” de la “mala” orientación, o las “consecuencias” de la “mala” orientación» (Silverman, 1997a, pág. 34), en la medida en que tales temas reflejan las concepciones de los problemas sociales tal como los reconocen los grupos profesionales o de la comunidad. No obstante, esto no obliga a abandonar la indagación causal en la evaluación cualitativa. La inda-

gación sobre cómo fabrican los profesionales de modo incesante unas formas y unas estructuras de servicio abre un mejor camino para comprender las causas.

Miles y Huberman (1994) sugieren una serie de consejos prácticos sobre diversas formas de abordar el análisis causal a través de la metodología cualitativa. Por ejemplo, los informantes locales tienen unos «mapas causales locales» en su cabeza y gran parte del trabajo de campo evaluativo implica obtener esos mapas y emplearlos en el desarrollo de unas redes causales más amplias. Las redes causales «no son probabilísticas, sino específicas y determinadas, basadas en la comprensión de los acontecimientos a lo largo del tiempo en el contexto local concreto, y vinculadas a una buena comprensión de cada variable» (Miles y Huberman, 1994, pág. 159). Los programas informáticos han incidido de forma importante en la posibilidad de contar con esos mapas.

Las redes causales existen en lo que respecta al caso individual. Miles y Huberman desarrollan también la forma de abordar la ordenación y la explicación mediante muestras de datos de diversos casos. Creen en la posibilidad de que la indagación cualitativa vaya más allá de los problemas de inferir a partir de la asociación («la palabra delatora de los investigadores cuantitativos»; 1994, pág. 222). Resumen el proceso de la forma siguiente:

La red causal entre diversos casos es un análisis comparativo de todos los casos de una muestra, empleando las variables que se calcule que son las más influyentes al exponer el criterio del resultado. Se observa cada medición de resultados y se examina, para cada caso, la corriente de variables que llevan a ese resultado o que lo «determinan». Las corrientes que son similares o idénticas en los distintos casos, y que difieren de forma más o menos constante de otras corrientes, se extraen y se interpretan (1994, pág. 228).

Señalan los pasos para conseguir este análisis, pero indican que es posible que no sea viable en muestras mayores o muy pequeñas.

EJEMPLO 9.4. MAPAS CAUSALES: MODELOS DEL PARTICIPANTE PARA UN CENTRO DE ACTIVIDADES RURAL

Shaw y sus colegas (1992) exponen una evaluación de un centro de actividades rural para personas con discapacidades para el aprendizaje. Observaron y entrevistaron a los participantes en el

	integración en el trabajo	vinculación con los cuidadores y los organismos de asistencia social	basadas en la comunidad y relaciones con el exterior; trabajo artesanal y destrezas para el trabajo del hogar
Personal	Especialistas en educación y necesidades especiales; voluntarios	Especialistas en trabajo social y en grupo; orientadores expertos	Especialistas en educación y asistencia social; voluntarios
Resultado	Rendimiento regular; prácticas laborales; normalización de los patrones de trabajo; aprendizaje de destrezas	Distinción poco clara entre el programa y los resultados	Amplia variedad de destrezas sociales; integración en las redes de la comunidad; cambio de actitudes por parte de los miembros de la comunidad externa

Fuente: Shaw, Williamson y Parry-Langdon (1992).

LA VALIDACIÓN DE LOS MIEMBROS Y LA TRIANGULACIÓN

Al hablar del análisis participativo, señalábamos la idea extrema de Tripp de que los datos son aquello que el entrevistado acepta de lo que se ha escrito sobre lo que se dijo. Si se le quita la excéntrica idea de la naturaleza de la labor evaluativa, puede parecer que Tripp presume la existencia de una verdad bastante evidente por sí misma: si queremos saber si hemos expresado adecuadamente las ideas, las actitudes y la conducta de quienes hemos observado o con quienes hemos hablado, debemos preguntárselo. Miles y Huberman, por ejemplo, reiteran a menudo su idea de que la táctica fundamental para comprobar la validez de nuestras conclusiones es «hacer que los informantes nos retroalimenten» (1994, pág. 165). Guba y Lincoln di-

cen también que remitir de nuevo los datos a las fuentes es la base para cumplir el criterio de validez. Además de las razones éticas a favor de este enfoque, también se puede encontrar apoyo a esta idea en la conciencia de las continuidades entre el razonamiento de sentido común de los miembros de la comunidad y el «razonamiento científico». Los autores que defienden esta postura a veces exigen que se compruebe la adecuación de las proposiciones científicas teniendo en cuenta si los miembros las entienden.

Bloor (1983) expone dos ejercicios de validación en los que preguntaba si los miembros reconocían, comprendían y aceptaban la descripción que él hacía. Realizó un estudio sobre las prácticas clínicas de varios especialistas en amigdalotomía y, a continuación, se le dio a cada uno de los especialistas un informe sobre sus prácticas clínicas, informes que fueron la base de las entrevistas que se les realizaron. Dice el autor que «los miembros reconocían mis explicaciones, pero no mostraban los mismos conocimientos de sentido común respecto a sus prácticas profesionales» (Bloor, 1983, pág. 160). El ejercicio se encontró con varios problemas:

- Algunos especialistas leyeron el informe sin excesivo interés.
- Había una sensación de artificialidad, en el sentido de que los puntos esenciales a veces los creaba la entrevista.
- «Una entrevista es una especie de conversación y, como tal, sigue las reglas implícitas de la conversación educada, en la que el desacuerdo manifiesto se reduce al mínimo» (1983, pág. 162). Por lo tanto, cuando el «desajuste de la realidad» aparece en la entrevista, conduce a un «trabajo de reparación conversacional», con el que ambas partes buscan el consenso.
- El ejercicio de validación de los miembros partía de la falsa premisa de suponer que las ideas de los cirujanos eran invariables, y no provisionales, contingentes y sometidas a cambio a lo largo del tiempo.

En resumen, el ejercicio de validación se vio obstaculizado por algunos problemas importantes; una lectura inadecuada por parte de algunos cirujanos, una falta enorme de interés, las influencias contextuales sobre las entrevistas, la naturaleza orientada al consenso del proyecto y el carácter esencialmente provisional de las reacciones de los distintos miembros (Bloor, 1983, pág. 164).

Emerson y Pollner (1988) exponen un ejercicio de validación de los miembros en el que surgieron problemas similares. Después de su

investigación de campo sobre los equipos de guardia de urgencias psiquiátricas, hablaban con un miembro del equipo durante el almuerzo y, posteriormente, hacían una exposición pública de sus resultados al Departamento de Salud Mental del condado. A un nivel muy básico, en muchos casos no estaba claro si su informe se confirmaba, se matizaba o se contradecía. Además, el contexto relacional en el que hablaban durante el almuerzo con un miembro del equipo con quien habían mantenido una relación muy estrecha, introducía «un elemento equívoco respecto a la medida en que las que parecen ser expresiones de acuerdo o desacuerdo están configuradas por el deseo del miembro de apoyar a un amigo» (1988, pág. 192). El contexto de la organización también configuraba los ejercicios. Evidentemente, los miembros pueden imaginar diversas formas de utilizar una investigación supuestamente «neutral» de manera que provoque unas consecuencias trascendentales para el personal de la organización. Por consiguiente, «las expresiones de apoyo o, para el caso, de hostilidad, se pueden construir según si las consecuencias del informe de la investigación fomentan uno u otro de los intereses en pugna» (1988, pág. 194).

Bloor se muestra escéptico sobre el valor de este tipo de ejercicios como parte de la validación. Su conclusión es que «las propias debilidades metodológicas que llevan a los sociólogos a buscar pruebas validadoras también se manifiestan en la generación de estas pruebas validadoras» (Bloor, 1997b, pág. 49). Más que como una validación de las pruebas, debemos considerar estos ejercicios como unas pruebas adicionales valiosas. «Los ejercicios de validación no son test, sino oportunidades para la elaboración reflexiva» (ibíd.) —en palabras de Emerson y Pollner, «menos como momentos que trascienden del enclave, y más como fuente de un material que se ha de seleccionar y considerar como parte del enclave» (1988, pág. 190).

Gran parte de todas estas ideas parecen sensatas. Pero me da la impresión de que las conclusiones de Bloor, Emerson y Pollner pueden ser excesivamente pesimistas e incluso huelen a elitismo académico. En lo que a la acusación de pesimismo se refiere, aunque podamos creer que los juicios de los profesionales no implican «ningún privilegio epistemológico, moral ni político sobre este tema», «el hecho de que el observador estuviera allí y/o de que los participantes crean que lo que se dice es verdad, son unas pruebas importantes de la validez de la explicación» (Hammersley, 1992, págs. 76, 52-53). Phillips señala que la cuestión está en la naturaleza de las conclusiones o de la explicación cuya credibilidad se quiere demostrar. Si es

«una explicación de las creencias de un individuo o de un grupo de sujetos [...] entonces el criterio correcto es si estos sujetos están de acuerdo o no en que el investigador haya registrado con exactitud sus creencias» (Phillips, 1987a, pág. 20). Si el propósito de la evaluación va más allá de comprender las perspectivas de los participantes, entonces la credibilidad es un sustituto inadecuado de la validez.

Por lo que se refiere a la acusación de elitismo, la negociación del acuerdo de los miembros sobre el significado se puede justificar con razones que no sean las estrictamente metodológicas. Lather estudió la oposición de los alumnos a la introducción de un currículo liberador y su conclusión era que «parece que las comprobaciones de los miembros producen el importante efecto de fomentar un creciente sentimiento de colaboración, en oposición a una validación negociada del nivel descriptivo» (1991, pág. 77). Los argumentos a favor de este tipo de comprobaciones de los miembros son, claro está, la única base con que puede trabajar el evaluador relativista. Por eso Smith concluye: «Se pueden realizar las comprobaciones por parte de los miembros porque es la única cosa moral y éticamente correcta que se puede hacer dadas las circunstancias, y no porque esas comprobaciones se vayan a traducir en un estudio más objetivo» (1992, pág. 103). La defensa de este tipo de ejercicios alegando razones morales no requiere una postura relativista, pero tampoco pierde su fuerza simplemente porque los resultados empíricos sean volátiles y contingentes.

Una lógica similar es la que rige en la aplicación de las ideas referentes a la *triangulación* a los datos de la evaluación. La idea de triangulación que más se suele utilizar se refiere a la combinación de dos o más estrategias de investigación diferentes en el estudio del mismo fenómeno empírico (Denzin, 1989). La idea se ha empleado de forma un tanto confusa en el discurso popular. Por un lado, a veces se entiende como una imagen para obtener una «fijación» geográfica sobre la verdadera «ubicación» de las pruebas. Por otro lado, muchas veces se entiende como una fuente de perspectivas y construcciones variables. La confusión es comprensible y recuerda los cambios de epistemología en las ciencias sociales. La triangulación como medio de validación basado en múltiples métodos se remonta a la obra de Campbell y Fiske sobre la validación convergente y discerniente, o lo que ellos llamaban unos enfoques de multirrasgos o de métodos múltiples. Estas obras están asociadas a una interpretación tradicional de la validez, como medio para separar la tendenciosidad de la realidad.

No hay duda de que una idea simplista de la triangulación corre el riesgo de que se la acuse de realismo ingenuo. Además favorece la

idea nada útil de que emplear múltiples métodos llevará a unas conclusiones más sólidas y mejor consensuadas, de forma fácil y aditiva. Un antídoto definitivo contra estos dos errores se puede conseguir en la lectura de una exposición antigua y clásica de los problemas de la conciliación entre los datos cuantitativos y cualitativos (Trend, 1979). Trend participó en una importante evaluación de unos experimentos sociales que realizaba el Departamento de Vivienda y Desarrollo Urbano (HUD) de Estados Unidos, diseñados para verificar la idea de utilizar la desgravación directa por la compra de vivienda, para ayudar a las familias de menores ingresos a conseguir una vivienda digna en el mercado abierto. Los datos cuantitativos de un estudio de casos sugerían que el programa producía mejoras importantes en la calidad de la vivienda de las personas que empleaban ese sistema. Sin embargo, los datos cualitativos y observacionales señalaban que el programa sería un fracaso. La mayor parte del informe de Trend lo ocupa la continua intención de dar una explicación verosímil. Su conclusión es que no siempre son compatibles los diferentes métodos de investigación: «No siempre es evidente la complementariedad. Limitarse a usar diferentes perspectivas, con la esperanza de que se validen entre sí no indica qué hay que hacer si las piezas no encajan» (1979, pág. 83).

Trend da un buen consejo sobre lo que hay que hacer en estas circunstancias: «Sugiero que demos oportunidad para que se manifiesten los diferentes puntos de vista y que suspendamos el rechazo inmediato de la información o las hipótesis que parecen que no se ajustan al punto de vista de la mayoría» (1979, pág. 84). En apoyo a esta idea, cita a Feyerabend, quien dice: «Parece que no es la actividad de resolver enigmas la que hace que aumenten nuestros conocimientos, sino la interacción activa de diversas ideas defendidas con fuerza» (citado en Trend, 1979, pág. 84).

LAS SÍNTESIS Y EL METAANÁLISIS

Las cuestiones del consenso y de la complementariedad de las pruebas de la evaluación plantean un tema más amplio del que se ha ocupado la literatura. ¿Hasta qué punto es posible y deseable sintetizar los juicios evaluativos en una afirmación general del valor y el mérito de un programa, una política o un proyecto determinados? Scriven siempre ha defendido que se hagan síntesis de las evaluaciones. Chelimsky, mientras estuvo en el Departamento de Contabilidad

General de Estados Unidos, se dedicó a fomentar la síntesis de estudios de evaluación múltiples. Cree esta autora que estos estudios, pese a los problemas de calidad de los datos, pueden ser un escudo contra «los dardos partidistas» más efectivo que los estudios únicos (Chelimsky, 1995). Hay aquí una distinción implícita. Chelimsky se centra en la *síntesis externa* (en el sentido de la integración de los estudios de investigación que se completaron como estudios singulares, pero que abordaban un tema común), mientras que Scriven insiste en la *síntesis interna*, en la que se evalúan los diferentes criterios para establecer el valor en un estudio determinado.

Los evaluadores cualitativos han sido tradicionalmente suspicaces. Stake, por ejemplo, dice que en su carrera profesional «cada vez he ido sustituyendo más la presunción realista por la duda constructivista». Esto explica por qué, «incluso como evaluador que cree que tenemos la responsabilidad de juzgar la calidad de un programa, soy reactivo a buscar afirmaciones sumarias y autorizadas» (Stake, 1991, pág. 81). Prefiere dejar que sea el público quien haga este tipo de juicios.

El debate ha tomado una dirección nueva después de que House defendiera los juicios sintetizadores. House está convencido de que «los criterios y los datos opuestos no impiden que los evaluadores, en la mayoría de los casos, lleguen a unos juicios después de considerar todas las cosas» (House, 1995, pág. 35). Es un tema sobre el que no hay sentencia aún, y habrá que esperar a más ejemplos de juicios de este tipo. Mi preocupación es doble. En primer lugar, parece que House subestima los problemas que plantea la resolución de conflictos en estos casos. El evaluador va a tomar parte de forma implícita. En segundo lugar, en estos sumarios se corre el riesgo de restar importancia o pasar por alto las contingencias contextuales locales. La base para transferir los conocimientos explicativos está en el contexto y la contingencia locales.

LA GENERALIZACIÓN

Los juicios y las conclusiones sumarias son un medio para conseguir una mayor aplicabilidad de las conclusiones de la evaluación. Los investigadores cualitativos hacen bien en recelar de la lógica convencional de la generalización basada en el muestreo de probabilidad. «El contexto y la historia locales [...] son las metas primordiales de la investigación cualitativa, y no la “generalizabilidad”» (Miller y Crabtree, 1994, pág. 348).

Nuestra respuesta a este escepticismo es que todo depende. En el capítulo 7 indicábamos de qué forma los diseños de la evaluación cualitativa, como los estudios de casos de enclaves múltiples, pueden tener en cuenta las exigencias de la generalización. También hemos hablado de cómo la explicación causal nos puede permitir extrapolar. Pero todo lo que pensemos en este campo debe tener en cuenta que la generalización exige una extrapolación que nunca se puede justificar del todo desde el punto de vista de la lógica.

El proceso de generalización está presente en todo aprendizaje y, si pensamos en cómo aprendemos en nuestra vida cotidiana, entenderemos mejor las cuestiones relativas a la generalización en la evaluación. Los enfoques reflexivos de la evaluación del profesional han tenido en cuenta este tema. La defensa que Eisner hace de la extrapolación a partir de su modelo de crítica educativa y de especialidad se basa casi enteramente en este tipo de argumentos. Distingue el autor el *generalizar*, en el sentido de la aplicación exacta a la situación B de lo que se ha descubierto en la situación A, del *transferir*, donde reconocemos que todas las situaciones son diferentes en alguna medida:

Como ninguna generalización se puede ajustar perfectamente a un contexto individual, siempre es necesaria la modificación. Esta modificación exige un juicio por parte de los profesionales inteligentes (Eisner, 1991a, pág. 212).

La expresión «transferencia del aprendizaje» representa exactamente esta distinción implícita.

Eisner plantea la interesante cuestión de dónde se centra la generalización. Quiere hacer viable la idea de la «generalización retrospectiva», que se produce cuando contemplamos bajo una nueva luz el pasado, nombrando las cosas de tal forma que podamos encontrar un nuevo sentido a toda una variedad de experiencias pasadas. «Encontramos un ajuste entre una afirmación general y nuestras historias personales. Las nuevas ideas pueden reconstruir nuestro pasado» (1991a, pág. 206).

Estas ideas encuentran un eco parcial en ideas más formales de cómo se produce la generalización en la indagación cualitativa. La evaluación cualitativa puede abordar de tres formas los problemas de la extrapolación: con la lógica de las encuestas; con la generalización analítica; y con la transferencia directa caso por caso a través de una descripción detallada y «gruesa» de ejemplos locales. Comentaremos brevemente cada uno de estos aspectos.

La lógica de la encuesta

Hablamos de la aplicación de la lógica de la encuesta a la extrapolación cualitativa sobre todo para subrayar sus limitaciones. En el capítulo 7 veíamos que la defensa de los estudios de casos de enclaves múltiples muchas veces ha estado en la base de la lógica de la probabilidad (Tizard y otros, 1975). Pero los problemas suelen ser más que los beneficios. Las conclusiones que se basan en probabilidades tienen una utilidad limitada en los casos concretos. Quizá podamos saber que una escuela, una intervención médica o un programa correccional han tenido una posibilidad «media» de tener éxito, pero esto no nos ayuda a valorar las contingencias particulares para una determinada decisión. Perdemos la «identidad del caso». Además, la situación típica a la que se enfrentan los profesionales, los padres o los gestores del programa no es una situación de generalización exacta, sino, para repetir la expresión de Eisner, de transferencia a unos enclaves o unas intervenciones en cierto grado diferentes. Cronbach, por supuesto, basa en este punto toda su defensa de la validez externa.

La generalización analítica

Nos hemos referido una o dos veces a la crítica dura y desmesurada que Atkinson y Delamont hacen de los estudios de casos en la investigación educativa. Sin embargo, donde sus críticas se agudizan es cuando se lamentan de que algunos representantes de la investigación mediante el estudio de casos sugieran que la generalización no interesa en los estudios cualitativos de instituciones particulares. Su respuesta es que tal idea se funda en una falsa comprensión de cómo puede tener lugar la generalización en esos casos (Atkinson y Delamont, 1993). Examinan el análisis comparativo, un enfoque que resulta familiar en los escritos interaccionistas y simbólicos. Es un método al que a veces se denomina inducción analítica. Consiste en utilizar la idea de un «problema genérico», que se aplica a una comparación explícita e implícita de las instituciones. El uso del método comparativo constante, incluida la identificación de los casos críticos y desviados, permite la identificación de las características comunes de las instituciones.

Este planteamiento del análisis ilustra que es fundamental una postura teórica sobre la evaluación. Además nos lleva a fijarnos en

los diversos enclaves y contextos de la evaluación. La obra de Goffman es un ejemplo clásico de este enfoque a través de sus conceptos formales de estigma e institución total. En general son muy pocas las evaluaciones que se realizan de este modo. Firestone ha demostrado que los temas de la generalización se pueden tratar de forma sistemática sin abandonar el modo analítico (Firestone, 1993). La idea de generalización naturalista e indirecta de Stake y Trumbull tiene importantes puntos de contacto en el acento que pone en las categorías analíticas generales, pero se asocia más con lo experiencial y privado, y supone un esquema moralmente consensual. Lo peor es que si los evaluadores cualitativos no están dispuestos a luchar con los conceptos y las teorías, «se limitan a sustituir una variedad de “conclusiones” ateóricas, basadas principalmente en la observación y la entrevista, por otra, basada principalmente en las puntuaciones obtenidas en unas determinadas pruebas» (Atkinson y Delamont, 1993, pág. 218).

Transferencia caso por caso y «descripción gruesa»

A lo largo del libro, hemos señalado implícitamente la importancia de la extrapolación basada en narrativas o relatos detallados del proceso, y no es necesario que insistamos más en ello. Como exhortación final bastará escuchar a Geertz cuando habla de la antropología. Lo que define a la antropología es el tipo de esfuerzo intelectual que es, una incursión detallada en la «descripción gruesa». Esto es lo que decía Geertz, en un artículo que se ha convertido en punto de referencia. Aunque aparentemente no le interesa la evaluación, es difícil leer un texto sobre la evaluación cualitativa sin encontrar esta expresión. Es la idea que el autor tiene de la relación entre lo oscuro y lo público, lo pequeño y lo de gran escala, que parece atraer por igual a evaluadores y etnógrafos. Su tesis es inspiradora más que didáctica y está construida en torno a una explicación de un trabajo de campo con judíos, bereberes y franceses en Marruecos en 1912, y una historia detallada de un robo de ovejas. El autor pone el acento en el carácter microscópico de las descripciones etnográficas: «Lo característico es que el antropólogo plantee [...] las interpretaciones más amplias [...] desde el punto de vista de la familiaridad cada vez mayor con cosas extremadamente pequeñas» (Geertz, 1973, pág. 21). Afronta las grandes realidades del Poder, la Fe, la Oposición, el Prestigio, el Estatus y el Amor, «pero las afronta en unos contextos lo suficientemente pequeños [...] para eliminar las mayúsculas» (1973, pág. 21).

Sobre el problema de cómo generalizar de lo local a lo general, Geertz rechaza la idea de que una determinada ciudad o un determinado lugar actúe como un microcosmos («una idea que seguramente sólo podría concebir alguien que llevara demasiado tiempo en el monte»; 1973, pág. 22). Al contrario, «la generalidad que logra alcanzar nace de la delicadeza de sus distinciones, y no de la extensión de sus abstracciones» (1973, pág. 25). En su frase más memorable, «los hechos pequeños hablan de grandes temas; los guiños, de epistemología; los robos de ovejas, de la revolución» (1973, pág. 23). Por consiguiente, «la tarea esencial de la construcción de una teoría [...] no es codificar unas regularidades abstractas, sino hacer posible la descripción gruesa; no generalizar entre los casos, sino generalizar dentro de ellos» (1973, pág. 26).

La tesis de la evaluación cualitativa nunca ha sido que sus pretensiones de generalizabilidad destaquen por su fuerza. Pero esta breve visión general de los enfoques analíticos, unida a lo que decíamos del diseño de la evaluación en el capítulo 7, demuestra que en cualquier caso no hay que evitar los métodos cualitativos por miedo a que sus pretensiones de una relevancia más amplia sean especialmente débiles. No es así.

DAR RELEVANCIA A LA EVALUACIÓN

¿Qué autoría hay que dar a la evaluación? ¿Quién debe hacerla y con qué voz o voces debe hablar? ¿Qué responsabilidad tiene el evaluador, si es que tiene alguna, sobre los usos directos o a largo plazo que se le puedan dar a la evaluación? En el capítulo 4 señalábamos la diferencia entre las posturas descriptiva y prescriptiva que adoptan los evaluadores cualitativos. Además, reconocíamos la diferencia entre quienes ponen el énfasis principal en el valor instrumental de la evaluación, y quienes consideran que ésta posee un valor de ilustración menos directo. También dedicábamos cierto espacio a considerar las posturas muy diferentes que los evaluadores adoptan en su papel político.

Pero cualquiera que sea la posición en estos temas, es probable que todos los evaluadores cualitativos estén influidos por lo que se ha descrito como la postura fundamentalmente apreciativa de la etnografía, como parte de lo que Finch llama la «tradición alternativa» dentro de la investigación de políticas inglesa. Evidentemente, los etnógrafos adoptan diversas posturas sobre el sentido en que la etno-

grafía es y puede ser «relevante». Hammersley, por ejemplo, cree que hay unas cuestiones importantes sobre la capacidad de la etnografía para producir unos resultados relevantes:

Surgen del hecho de que las situaciones que estudian los etnógrafos raramente tienen una relevancia intrínseca. Al contrario, hay que hacer relevantes las conclusiones apelando a la generalización empírica o a la inferencia teórica (Hammersley, 1992, pág. 77).

La comprensible respuesta de los evaluadores debería ser afirmar la relevancia intrínseca de algunos (no todos, desde luego) de los problemas que tratan. Pero, como ya hemos visto en el capítulo 6, las dudas de Hammersley permanecen. Al hablar del movimiento del profesor como investigador, señala algunas razones para creer que la investigación no puede resolver de forma rutinaria los problemas del profesor (Hammersley, 1993a).

En mi opinión, Hammersley está demasiado influido por su preocupación por advertir contra ciertos procedimientos que teme que van a socavar la calidad y la credibilidad de la investigación. Los etnógrafos pueden adoptar de forma convincente una postura más optimista. En los últimos párrafos del capítulo 6, por ejemplo, decíamos que Bloor y McKeganey demuestran la relevancia potencial de la etnografía. Y Holstein y Gubrium, cuando explican una entrevista a una mujer de 88 años, ofrecen un ejemplo revelador de cómo sus procesos de entrevista activa pueden producir a la vez una investigación rigurosa y unos beneficios personales para el entrevistado (1995, págs. 59-65).

¿Coautoría?

¿En la redacción del informe final, los evaluadores han de colaborar con quienes participaron en el estudio? ¿Qué forma ha de tener ese informe? ¿Cómo han de manejar los evaluadores el potencial y las dificultades de la defensa en relación con el informe?

Recordemos que Scriven considera que la autoría participativa del informe final constituye la práctica de evaluación más pobre de todas las posibles (1997, pág. 486). Volviendo a la idea de coautoría y negociación de Tripp, tal vez nos sorprenda que coincida con quienes piensan que la coautoría no tiene nada de rigurosa. Decíamos que aboga por la coautoría en la producción de datos. Pero a partir de ahí

pasa a distinguir claramente entre los datos y el informe. Habla de sus propias investigaciones, en las que «incitaba» a muchos de sus entrevistados a que aceptaran «unas afirmaciones sinceras e incluso provocativas», por ejemplo cuando decía que se ponían en peligro las relaciones de trabajo en la institución (Tripp, 1983, pág. 38). Sin embargo, insiste en que cuando los entrevistados convienen en unas afirmaciones, «son responsables personalmente ante sí mismos, de forma que se descarta que el efecto se pueda moderar de alguna manera» (ibíd.). Pero no sólo esto, sino que además piensa «que el autor del informe de una evaluación debe intentar mantener el control sobre la interpretación hermenéutica» (1983, pág. 43), de modo que el investigador tiene unos derechos de interpretación que al lector se le niegan. Es fácil imaginar las reacciones de quienes participaban en sus estudios ante tal *mélange* de imágenes democráticas y autoritarias sobre los conocimientos y las destrezas de evaluación.

Los problemas a que nos referíamos cuando hablábamos de la validación de los miembros se plantean una vez más en relación con la coautoría de los informes, además de otras dificultades más prácticas.

EJEMPLO 9.5. LOS PRINCIPIOS DEMOCRÁTICOS EN LA PRÁCTICA

Simon describe el proyecto SAFARI (siglas inglesas de Éxito y Fracaso e Innovaciones Recientes), que era una evaluación del impacto de cuatro proyectos curriculares en el nivel de educación secundaria en Gran Bretaña. Eran proyectos de tecnología, geografía, humanidades y ciencias.

En una reunión con jefes de departamento, la autora insistía en tres puntos referentes al espíritu participativo del proyecto:

En primer lugar, las aspiraciones participativas de la investigación. Segundo, la oportunidad que el control compartido les daba de configurar el estudio y determinar el proceso y la amplitud de su difusión dentro y fuera de la escuela. Y en tercer lugar, la responsabilidad que quería que asumieran de identificar mis tendencias personales y profesionales [...] Era «su» estudio tanto como el mío (Simon, 1987, págs. 96-97).

La segunda observación tiene una relevancia particular para las prácticas de la coautoría. La autora completó unas breves evaluaciones etnográficas con visitas de tres días a cada uno de los centros educativos, y luego empleó uno o dos días más en aclarar los datos con las

personas implicadas. Todo esto se vio superado por un problema importante referente a la confidencialidad de los datos, lo cual indica que «los procesos de control interno de datos que no son familiares, con los que se pretende proteger a los participantes, se pueden ver superados fácilmente por los procesos tradicionales de las quejas» (1987, pág. 110). Había en esto aspectos positivos, en el sentido de que subía el «nivel de realidad» de la indagación, pero demuestra que la evaluación receptiva puede escapar fácilmente del «control logístico».

El tiempo necesario para llevar a cabo la evaluación participativa se convirtió en una cuestión importante. Simon hizo al menos tres borradores de cada uno de los cinco informes, y al final tuvo que abandonar el proceso: «Había empezado algo que no podía terminar» (1987, pág. 125).

Presentación de las conclusiones

En los debates sobre la presentación de las conclusiones de la evaluación se suelen ocultar planes diversos. A Finch, por ejemplo, le preocupan las ideas negativas de los responsables de las políticas y de quienes toman las decisiones referentes a la investigación cualitativa. En consecuencia, sostiene que los evaluadores deben dar la suficiente información sobre la metodología del estudio para que quien formule las políticas pueda valorarla, y para demostrar que en todo momento se ha mantenido el control sobre los métodos. Si la evaluación se realiza siguiendo las líneas del diseño y el análisis participativos, se presentará de forma coherente poniendo énfasis en el proceso del trabajo tanto como en los resultados. Las versiones de la evaluación en las que priman los interesados adoptarán diversas formas de informe y de retroalimentación. Por lo tanto, estas diferentes posturas sobre la presentación de los informes son contingentes. El verdadero tema por tratar es el propósito evaluativo y la exigencia de que se presente la evaluación de forma que se refuerce ese propósito.

Pero se puede decir algo más. Son muchas las obras que ven la indagación cualitativa como una forma de narración de historias. Miller y Crabtree, en su exposición de la investigación clínica, dicen que compartir los conocimientos conlleva «contar unas historias que sean metodológica, retórica y clínicamente convincentes» (1994, pág. 348).

La investigación clínica cualitativa es convincente si los métodos son los adecuados para el tema en cuestión y si la relación del investi-

gador con los informantes, los datos y el público se aborda claramente; si el público se reconoce en las conclusiones; y si a los participantes clínicos les importan el tema y los resultados (1994, pág. 349).

EJEMPLO 9.6. LEE CRONBACH SOBRE LAS HISTORIAS CONTABLES

El evaluador se encuentra ante una suave paradoja. «Toda evaluación pretende reducir la realidad a una historia que se pueda contar», pero «el estudio exhaustivo de un problema social hace que parezca más complicado» (Cronbach y otros, 1980, pág. 184). La solución que los autores dan a esta paradoja está en el aforismo que dice que el examen comprensivo no equivale al informe exhaustivo. «Cuando se produce una avalancha de palabras y de tablas, todos los que se encuentran a su paso se apartan» (1980, pág. 186).

El principal criterio es la medida en que las personas relevantes aprenden a partir de lo que el evaluador les comunica. Por consiguiente, el evaluador debe buscar a lo largo de la evaluación la permanente oportunidad de comunicarse con la comunidad que configura las políticas. Piensan los autores que «gran parte de la comunicación más importante es informal, no toda ella es deliberada, y algunos de los mayores efectos son indirectos» (1980, pág. 174). Sus recomendaciones son:

- Estar presente.
- Hablar poco y a menudo.
- Contar historias, estar siempre preparado con una reserva de anécdotas referentes a la evaluación.
- Hablar con las fuentes del director.
- Utilizar múltiples modelos de presentación.
- Ofrecer justificaciones públicamente defendibles de cualquier cambio que se recomiende en el programa. Serán unos argumentos muy distintos de los científicos.

Cronbach se opone con energía a seguir hasta que se cuente con todos los datos y hasta que las conclusiones sean firmes. La influencia y la exactitud estarán en constante tensión y, en caso de duda, siempre hay que optar por la influencia. Las visiones generales en directo, informales y rápidas, la respuesta a las preguntas, el uso de películas y de grabaciones en audio y las apariciones personales son el material de que se abastece la influencia. Por lo tanto, el informe final debe actuar como un documento de archivo:

La impotencia que acompaña al retraso [...] puede ser un coste mayor que las consecuencias de un juicio equivocado. El proceso político está acostumbrado a la defensa enérgica [...] [y] no la va a borrar la afirmación mal considerada ni siquiera de un evaluador (1980, págs. 179-180).

La defensa

Sin embargo, podemos explicar nuestra postura respecto a la defensa o propugnación: «A corto plazo, la buena investigación social se recibirá a menudo como la traición de una u otra parte de una determinada controversia» (Rossi, 1987, pág. 79). La escarmentada conclusión de Rossi, un tanto burlona, es que «ninguna buena investigación social se libra del castigo».

Incluso los evaluadores comprometidos con la defensa pueden tener sus dudas. Stake indica que existen algunas orientaciones para distinguir entre buenas y malas defensas:

Creo que nuestra postura debe ser la de oponerse a la defensa, minimizar la defensa, aunque siempre esté presente, ayudar a reconocer las defensas que tienen un valor y las que no lo tienen, pero oponernos a compartirlas (1997, pág. 474).

Si nuestra postura es la del multipartidista, no es excusa para la inactividad. El multipartidismo es una postura activista. Rebasa la complacencia del pluralismo liberal que se manifiesta en algunas formas de la evaluación receptiva e interpretativa. Eraut (1984), por ejemplo, pide lo que él denomina una evaluación divergente que permita que el evaluador introduzca unos valores a los que los entrevistados no prestan atención, aunque nos les dé ninguna prioridad especial.

Yo me inclino por una evaluación reformista, en la que el evaluador ha de buscar los datos que con mayor probabilidad respalden los programas que benefician a los desposeídos. Reconozco la fuerte impresión de disonancia que puede producir la defensa de una metodología evaluativa rigurosa en un momento, y la defensa reformista y partidista basada en las pruebas, en otro. Pero de momento adopto esta postura y, además, con la convicción de que, reiterando las palabras de Jennifer Greene:

Sin una postura comprometida y responsable como la de la ciudadanía científica, me temo que la evaluación cualitativa quedará marginada de los debates sobre los acuciantes temas públicos [...] Ya no podemos ocultar nuestra condición de ciudadanos tras nuestras subjetividades científicas. Debemos convertirnos en ciudadanos científicos (1996, págs. 286-287).

APÉNDICE

EJERCICIOS

Los ejercicios de las páginas siguientes nos darán la oportunidad de explorar los temas y la metodología de los que se habla en este libro. Algunos están pensados para consolidar lo aprendido. Otros son más participativos y nos invitan a emprender tareas relacionadas con la evaluación, o a realizar un juego de simulación. He vinculado cada uno de los ejercicios con el capítulo con el que está más relacionado, pero hay que realizarlos siempre de forma flexible y desde el punto de vista del aprendizaje propio. Algunos se hacen mejor como ejercicios individuales y otros como ejercicios de grupo, pero también aquí hay que ser flexibles.

CAPÍTULO 1. ¿QUÉ ES UNA EVALUACIÓN CUALITATIVA?

Este ejercicio de observación tiene dos fases. Cada una requerirá más o menos una hora. Funciona mejor si la primera fase se empieza una semana antes que la segunda. A los participantes les suele gustar este ejercicio y consideran que merecen la pena el tiempo y el esfuerzo que hay que invertir para completarlo. Naturalmente, se puede adaptar a las exigencias de un curso convencional sobre metodología.

Fase 1

Durante no más de quince minutos observemos una situación en la que normalmente participemos. No importa qué tipo de situación sea. Puede ser en la familia, en el lugar de trabajo o entre las personas con quienes nos relacionamos en otros lugares.

Antes de iniciar la observación, dediquemos unos pocos minutos a pensar sobre esa situación o ese enclave y decidir dónde podemos

centrar la observación. Es imposible no ser selectivo y, dado el corto período de observación, deberemos tener algunas ideas claras sobre el tipo de cosas que podamos buscar. Sin embargo, no es aconsejable decidir un único objetivo de la observación, ya que deberemos responder a aspectos nuevos que se plantearán en la situación.

Tendremos que decidir de antemano cómo vamos a registrar la observación. Podemos hacerlo mientras observamos o podemos tomar notas, para pasarlas a limpio lo antes posible una vez finalizada la observación. También habrá que decidir si revelamos lo que estamos haciendo a los demás implicados en esa situación.

Una vez concluida la observación y cuanto antes, escribiremos tres o cuatro párrafos en los que generaremos todas las explicaciones que podamos de los sucesos que hayamos descrito. Dedicaremos unos 45 minutos a analizar así nuestra observación.

Fase 2

Leamos de nuevo las páginas sobre las características de la evaluación *cualitativa* del capítulo 1. Identifiquemos qué características de la indagación cualitativa se reflejan en la descripción y el análisis que hayamos hecho. Si compartimos el trabajo de observación con otras personas, esta fase nos llevará más o menos una hora.

CAPÍTULO 2. PERSUASIONES Y PERSUASIVOS

Leamos de nuevo la lista de preguntas con que termina el capítulo 2.

1. Copiemos la lista y señalemos qué teórico o teóricos han ayudado a centrar la atención en cada uno de los temas.
2. ¿Cuáles son, en nuestra opinión, los tres o cuatro temas de la lista más fundamentales para el desarrollo de una evaluación cualitativa en la situación que nos sea más familiar? Si trabajamos en un grupo, deberemos comparar nuestras listas de prioridades para las diferentes situaciones.

CAPÍTULO 3. PARADIGMAS Y PREGUNTAS DE LA EVALUACIÓN

Escribamos notas breves sobre una situación que nos sea familiar. Puede ser una innovación en el currículo o en la asistencia social, la aplicación de la legislación en nuestro lugar de trabajo, un programa de vivienda, un nuevo programa de evaluación universitaria, etc.

Imaginemos que hay que evaluar esa situación. Escribamos las tres preguntas de evaluación que más preocuparían a cada una de las siguientes personas:

- Un pospositivista realista.
- Un evaluador crítico.
- Un evaluador constructivista.

CAPÍTULO 4. PARTIDISMO Y EVALUACIÓN

Leamos de nuevo las páginas del epígrafe «Validez y valores» del capítulo 4.

¿Cuál es nuestra postura respecto a los temas del partidismo? ¿Somos sobre todo evaluadores multipartidistas, reformistas o abiertamente ideológicos?

Hagamos de abogado del diablo con nuestras propias simpatías y preguntémonos cuáles son los *defectos* principales de la postura que apoyamos.

En este ejercicio se conseguirá una dimensión útil si somos capaces de crear un foro en el que defendamos los resultados de nuestras reflexiones ante un pequeño grupo de alumnos o evaluadores compañeros nuestros.

CAPÍTULO 5. VALORAR LAS EVALUACIONES DE PROGRAMA

En el capítulo 5 hemos dicho que los evaluadores cualitativos:

1. Considerarán el *proceso* de las actividades y la aplicación del programa antes de considerar sus resultados.
2. Desconfiarán de la aplicación y la evaluación del programa que dependan en gran medida de supuestos de una *toma de decisiones racional*.

3. No deben ignorar la importancia que tiene *teorizar* sobre los programas.
4. Serán menos entusiastas sobre el valor de los modelos de evaluación de *rendición de cuentas*.
5. Incluirán una diversidad de posturas acerca del lugar que ocupa la evaluación de programas en la contribución al *cambio social*.

Escojamos un ejemplo preparado o publicado de evaluación que incluya una metodología cualitativa. Averigüemos, con cualquier medio que sea apropiado (revisión de estudios, debate en grupo, conversaciones con los interesados, etc.), hasta qué punto esa evaluación refleja lo que decíamos sobre esos cinco criterios en la primera mitad del capítulo 5.

CAPÍTULO 5. EL USO DE LA INFORMACIÓN

Carol Weiss y sus colegas entrevistaron a 155 personas que ocupaban cargos muy importantes en organismos de salud de ámbito federal, estatal y local. Entre las preguntas que hacían estaban las siguientes:

1. ¿Usa usted conscientemente los resultados de la investigación de las ciencias sociales para tomar las decisiones en su trabajo?
2. ¿De qué forma emplea usted la investigación de las ciencias sociales en su trabajo?
3. ¿Busca usted información de investigaciones cuando considera las alternativas políticas o de programa?
4. ¿En qué circunstancias busca usted una investigación? (Weiss, 1980).

Las dos primeras preguntas se refieren al *uso consciente* de la investigación. Las otras dos, a la *búsqueda activa*.

Programemos una entrevista con un directivo importante (que conozcamos o no) y hagámosle esta serie de preguntas. Comparemos los resultados de esta entrevista con lo que decíamos en la segunda mitad del capítulo 5.

CAPÍTULO 6. LA EVALUACIÓN REFLEXIVA MEDIANTE LA ENTREVISTA A IGUALES

Para realizar este ejercicio deberemos ser profesores, trabajadores sociales, policías, sanitarios o algo similar. El ejercicio requiere que trabajemos con otra persona. Observemos que se supone que nosotros seremos el entrevistado, no el entrevistador. Sin embargo, es posible actuar como el entrevistador. En este caso, prepararemos una entrevista con alguien que desempeñe algún trabajo de prestación de servicios. Nosotros formaremos parte de una entrevista entre iguales, que durará entre 30 y 75 minutos.

El entrevistador debe disponer de una grabadora de buena calidad. Asegurémonos de que no se nos interrumpa durante la entrevista. Debemos dedicar unos pocos minutos (no más) a pensar en las preguntas antes de empezar, en particular para decidir de qué dos trabajos vamos a hablar. Este ejercicio funciona mucho mejor si se puede trabajar con la transcripción de la entrevista. Hay que tener en cuenta que la transcripción nos costará ocho veces más tiempo que la entrevista. Por ejemplo, la transcripción de una entrevista de 40 minutos nos llevará unas cinco horas. Pero merece la pena. Nosotros mismos podemos «dar el parte», pero necesitaremos a alguien que reflexione sobre él. Todos los participantes deben estar de acuerdo en que la entrevista y la cinta no se utilizarán con ningún otro fin que no sean los acordados de antemano. Este ejercicio se puede realizar en conexión con las preguntas de que hablábamos en el capítulo 8.

He hablado en el libro de algunos resultados de estudios de este tipo. En la investigación que se expone en Shaw y Shaw (1997a, 1997b) se hizo una secuencia de preguntas similar.

Las preguntas de la entrevista

[Nota para el entrevistador: Esta entrevista es abierta y sólo está vagamente estructurada. Habrá otros apuntes que querrá hacer el entrevistador, y la secuencia de la entrevista variará de algún modo. Sin embargo, la entrevista debe beneficiar a la persona que habla y el entrevistador ha de evitar por todos los medios buscar su propio interés.]

1. ¿Cuál es su reacción cuando oye hablar de evaluación? Piense un momento y haga una asociación de las palabras que se le ocurran.

2. Se insiste cada vez más en que las personas que trabajan en lo que usted trabaja deben evaluar lo que hacen. ¿Qué le parece la idea? ¿Cree que la evaluación es realmente importante, o no? ¿Intenta usted evaluar su propia práctica?

En la siguiente parte de la entrevista voy a pedir al lector que hable de dos ejemplos recientes de su trabajo. Quisiera que hablara de un ejemplo de su trabajo en el que crea que hacía bien las cosas, y un ejemplo en el que crea que no las hacía tan bien. Por favor, dediquemos un momento a decidir en qué orden queremos hablar de estos ejemplos.

3. ¿Podría empezar a describir brevemente su trabajo, y sobre todo lo que iba bien y lo que no iba tan bien?
4. ¿Cómo sabía si hacía eso bien o no tan bien?
5. ¿En qué pruebas se basaba realmente en ese caso? (Por favor, ajústese al ejemplo.)
6. Si pudiéramos hablar de este ejemplo de su trabajo con los usuarios, ¿cree que pensarían más o menos lo mismo que usted? ¿Cree que emplearían los mismos criterios para decidir si el trabajo había ido bien o no, o usarían otros distintos?
7. ¿Participaron de algún modo en la evaluación los usuarios, los gestores, los colegas, los cuidadores u otros profesionales? De ser así, ¿cree usted que habrían empleado los mismos criterios para decidir si el trabajo fue bien o no, o utilizarían otros distintos?
8. ¿Cree que la vara que empleó para juzgar si hacía bien o no las cosas era la adecuada para ese caso particular?

Repetiremos las mismas preguntas respecto al segundo ejemplo de trabajo bien o mal hecho.

9. ¿Se deduce algo más de estos casos? Por ejemplo, ¿ilustran alguna cosa relacionada con sus actitudes o sus ideas sobre la evaluación?

Preguntemos si hay algo de lo que hayan dicho que deseen aclarar.

Una vez terminada la entrevista, pensemos en las implicaciones que los resultados tienen para los cuatro tipos de evaluación del profesional de que hablábamos en el capítulo 4.

CAPÍTULO 7. EL DISEÑO DE LA EVALUACIÓN. UN JUEGO DE SIMULACIÓN:
«JÓVENES Y PROSTITUCIÓN»

Introducción y esbozo general

Hay que dar una copia de lo que sigue a los facilitadores y explicarlo verbalmente a los participantes.

En este ejercicio va a dedicar usted la mayor parte del tiempo a un juego de simulación, como miembro de un Grupo Consultivo que se ha reunido tres veces en los últimos seis meses para acordar una propuesta de investigación de evaluación. El ejercicio versa sobre el *diseño* de la evaluación, por consiguiente nos centraremos sobre todo en el *proceso* de la investigación, más que en sus resultados finales.

El ejercicio durará una hora. Se necesitan ocho «jugadores», con un mínimo de siete y un máximo de diez. Además se necesitan uno o dos facilitadores.

A lo largo del ejercicio hay que recordar que el principal propósito no es estudiar la dinámica del grupo ni evaluar la actuación de nadie en el papel que represente. El propósito es descubrir formas de gestionar los procesos de investigación de manera que ésta resulte aplicable, relevante y participativa.

[*Nota:* Las competencias de los papeles que se representan no se refieren a organismos u organizaciones reales, pero podemos decidir darle un toque de realismo y acordar unos nombres para los organismos, por ejemplo para las diversas instituciones infantiles.]

En las primeras reuniones las cosas han ido muy bien y existe un buen grado de consenso sobre el tipo de investigación que se necesita. Hay buena confianza entre los miembros del grupo. Todos estamos comprometidos en producir una investigación relevante y que resulte aplicable, y también en acordar un proceso que sea muy participativo. Hemos sabido colaborar pese a los diferentes puntos de vista y planes de los distintos organismos representados. El propósito de la reunión simulada del Grupo Consultivo es establecer un acuerdo sobre la forma concreta en que el proyecto va a alcanzar esos objetivos. Es el único punto del día.

Fase 1. Instrucciones generales

Distribuimos la hoja de instrucciones generales (véanse las páginas 281-282) y dedicamos *5 minutos* a familiarizarnos con su contenido. El facilitador será el encargado de aclarar cualquier duda que pueda haber en esta fase, y se asegurará de que la simulación cumple con el horario establecido. Todavía no hemos empezado el juego de simulación.

Los facilitadores pueden decidir distribuir las instrucciones antes de la simulación para que se puedan estudiar mejor.

Fase 2. Instrucciones individuales

Cada uno de los miembros tomará uno de los sobres de *instrucciones individuales* (págs. 283-287). Éste será nuestro papel, que tendrá sus correspondientes instrucciones. Si sólo hay siete «jugadores», antes de abrir los sobres retiraremos el número tres. Si hay más de ocho «jugadores», escogeremos al azar unos jugadores que actuarán de observadores, y deberán observar la reunión y ayudar a facilitar la retroalimentación en la Fase 4.

Durante el ejercicio se compartirá el contenido de los sobres, pero sólo hay que hacerlo como una parte planificada de nuestra participación en el juego de simulación. *No* empecemos preguntando: «¿Qué hay en *tu* sobre?».

Dedicaremos *5 minutos* a meternos en el papel. Esto supondrá una lectura atenta de nuestras instrucciones, preparar la forma en que vayamos a representar nuestro papel, recordando, mientras lo hagamos, la información que se nos ha facilitado sobre el tono y el contenido de las reuniones anteriores del grupo. Cuando consideremos que estamos lo suficientemente familiarizados con nuestro personaje, deberemos dedicar *unos pocos minutos* a hablar informalmente con el resto del grupo. Ahora estamos ya en nuestro papel y el propósito de esta conversación informal es reanudar el contacto con los compañeros del grupo después de las semanas transcurridas desde la última reunión, y asegurarnos de que sabemos quiénes asisten a la reunión.

Fase 3. Reunión del Grupo Consultivo

El presidente abrirá la sesión. La reunión durará aproximadamente *30 minutos*.

[*Nota:* Hay que procurar interpretar bien el papel que se nos haya asignado. Si somos un observador, asegurémonos de que no «contaminamos» la simulación. Así podría ocurrir, por ejemplo, si nos sentáramos demasiado cerca de donde se desarrolla el juego, si nos riéramos ante comentarios graciosos, si mantuviéramos una conversación no verbal con los participantes y, naturalmente, si habláramos mientras se representa el juego.]

Fase 4. La representación

Los dos grupos más pequeños se reúnen y dedican *15 o 20 minutos* a compartir las decisiones que se hayan tomado en cada grupo. Volveremos al tercer párrafo de esta introducción para recordar el propósito del ejercicio.

Deberíamos:

- Asegurarnos de que nos dirigimos a las personas pensando que no son actores, sino que realmente ocupan el cargo que se les ha asignado.
- En primer lugar, cada uno debería anotar las decisiones que crea que se hayan tomado en la reunión.
- Luego hablaremos de si estaban claras estas decisiones y si estábamos de acuerdo con ellas. De no ser así, ¿por qué y cuáles son las consecuencias?
- Reflexionemos sobre si «expresamos» bien lo que el papel que teníamos asignado suponía. La cuestión aquí es que si los organismos clave no manifiestan las opiniones de sus grupos de interés *en* el grupo, luego las van a manifestar dentro de su propio ámbito, con lo que no es probable que se alcance un auténtico consenso respecto al proyecto. Es en parte una función del presidente, pero también, sin duda, de cada persona que representa su papel.
- Identifiquemos el potencial y los obstáculos de planificar una investigación que resulte aplicable, relevante y participativa.

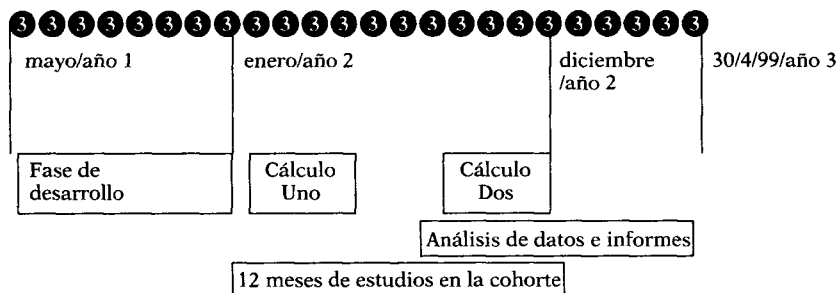
La Fase 4 la dirigirá el facilitador o los facilitadores.

Reunión del Grupo Consultivo

En tres reuniones anteriores del grupo se acordó apoyar la solicitud de una beca de investigación presentada por un grupo de personas de la universidad. Esta investigación:

1. Calculará el índice de prostitución infantil y juvenil en una gran ciudad.
2. Identificará cómo se inician los niños y los jóvenes en la prostitución y, en especial, cómo salen de ella.
3. Hará recomendaciones para la política y la práctica.

La investigación combinará métodos cuantitativos y cualitativos, y durará dos años. El diseño del estudio se resume en el siguiente *Esbozo del Programa*:



Nos reunimos en las dependencias de la Federación de Indigentes. Entre los organismos representados en esta reunión o en otras anteriores están:

- La Federación de Indigentes, una red federal de las principales organizaciones dedicadas a las personas sin techo de la ciudad.
- Varias grandes organizaciones que atienden a personas indigentes y que también hacen campaña a favor de los intereses de estas personas.
- Los principales responsables de cuatro importantes organizaciones infantiles distintas.
- Los líderes de proyectos en que se trabaja con determinados grupos, por ejemplo homosexuales jóvenes, jóvenes sin techo, muje-

res jóvenes, personas seropositivas con problemas de vivienda, personas relacionadas con la prostitución.

- La Asociación de Responsables de Trabajo Social (ACOSW). Un representante de esta asociación, que preside el Comité Infantil y Familiar.
- El comisario de policía de la zona.
- Personal médico que trabaja en el campo de la prostitución.
- Dos investigadores de la universidad.

A la reunión asistirán el Director de la Federación de Indigentes (que preside el grupo), los responsables de un proyecto de calle para mujeres jóvenes y de otro para hombres jóvenes, un investigador de un organismo infantil nacional, el director de un gran organismo para personas sin techo, un funcionario en calidad de presidente de un importante comité de la Asociación de Directores de Organizaciones de Asistencia Social, y uno o ambos investigadores. Los otros representantes han excusado su asistencia.

Instrucciones para cada personaje

Los facilitadores entregarán las instrucciones individuales en sobres cerrados a cada uno de los «jugadores».

1. Director de la Federación de Indigentes

Está usted de acuerdo con el enfoque que los investigadores dan al estudio y está dispuesto a actuar de coordinador entre los organismos que se ocupan de los sin techo. Le interesa la relación entre indigencia y prostitución.

El organismo que usted dirige es una red federal que actúa en nombre de una gran cantidad de entidades similares dedicadas a los indigentes. No prestan ustedes un servicio directo.

Usted es el Presidente del grupo. Ha recibido la notificación de que no podían asistir a la reunión el organismo infantil X (que adjuntaba una nota en la que decía que confiaba en que se diseñara una política sólida sobre la investigación participativa), el organismo infantil Y, el equipo médico, Acción Municipal en favor de la Salud Sexual (CASH) y Nuevos en la Ciudad (uno de los miembros activos de la Federación de Indigentes). Le interesa a usted asegurar que todos los organismos, incluidos los que han excusado su asistencia, se sientan satisfechos de lo que se decida, aunque no es un tema que le preocupe mucho.

2. *Investigador de la Universidad (1)*

Usted representa el papel principal y será quien dispondrá de la mayor cantidad de la beca si el proyecto tiene éxito. Apoya usted sin reservas la idea de la investigación participativa, aunque su universidad tiene la sede en otra ciudad y le preocupa un poco que todo el complejo aparato de dirección del proyecto pueda poner en peligro su éxito. Pero está dispuesto a comprometerse a contar con el consenso de todos los usuarios.

Además quiere disponer de un *panel de especialistas* de la universidad que puedan hacer aportaciones multidisciplinares en los puntos clave y confía en que se pueda conseguir con todo lo demás que se acuerde.

Tiene usted sus propias ideas, pero es muy consciente de que no puede llevar adelante la investigación si no cuenta con el apoyo de estos organismos clave. Apoyará usted al presidente para facilitar unos buenos resultados.

3. *Investigador de la Universidad (2)*

No le entusiasma mucho la investigación participativa. Confía más en la parte cuantitativa de la investigación y quiere que el diseño sea «de calidad». Pero espera que no tenga que discutir este asunto.

En anteriores reuniones ha oído que a los representantes de algunos organismos les preocupa el peligro de crear entre los jóvenes unas expectativas que luego, cuando el proyecto termine, no se puedan cumplir. También a usted le preocupa, y confía en que el grupo incluirá una *política sobre la ética de la investigación* que proteja a los jóvenes implicados. Ha decidido que la va a apoyar.

4. *Organización de Indigentes de Middlepoint*

Es usted la persona responsable de una gran organización de indigentes que se dedica a prestar servicios y a organizar campañas en favor de esas personas. Su organismo es uno de los miembros dirigentes de la Federación de Indigentes y piensa que el Director de la Federación es uno de sus aliados clave.

Tiene una experiencia muy positiva de la investigación participativa y quiere que ésta sea una parte sustancial del proyecto, que debe evitar el formulismo. Está impresionado por la cohesión que ha demostrado el grupo al discutir temas polémicos. Sin embargo, tiene la

ligera sospecha de que el representante de la Asociación de Responsables de Trabajo Social (ACOSW) quiere impedir a toda costa que se critique a los departamentos de asistencia social por no proteger a los jóvenes que ellos «atienden» para evitar que sufran la explotación de grupos de pederastas.

Quiere ver usted un *grupo consultivo de jóvenes* importante, que trabaje como grupo centrado a lo largo de todo el proyecto, aunque no ha pensado en los detalles de su posible creación. Defenderá esta idea en la reunión.

5. *Proyecto de calle para hombres jóvenes*

Es usted un trabajador social de un organismo de calle bien asentado que dirige una serie de trabajos para fomentar las solicitudes de prestaciones a las que se tiene derecho entre jóvenes varones homosexuales, algunos de los cuales tienen problemas con las drogas y la mayoría de ellos son seropositivos y se dedican al negocio del sexo. Le entusiasma el proyecto y cree que sintoniza con los investigadores. De hecho, confía en conseguir un puesto de investigador en el proyecto (es usted licenciado y ha realizado algunas investigaciones a pequeña escala que el examinador externo elogió mucho).

Cree que algunas de las otras organizaciones de voluntarios que trabajan en la ciudad no se han empleado como podrían en investigaciones que resulten aplicables y relevantes. Piensa que su propia organización podría desempeñar un importante papel para garantizar un proyecto auténticamente participativo y podría ofrecer sus dependencias para las reuniones del grupo. Está convencido de que hay que pensar que los jóvenes tienen un papel directo como coinvestigadores.

En reuniones anteriores ha expresado su preocupación por el peligro de que se puedan crear entre los jóvenes unas expectativas que luego no se cumplan, cuando concluya el proyecto. Recuerda que el segundo investigador le apoyaba, pero no está seguro de si va a asistir a la reunión. Apoyará que se elabore una *política sobre la ética de la investigación* que proteja a los jóvenes implicados.

6. *Investigador de la Organización Infantil Nacional A*

Su organización desea tomar la iniciativa entre las organizaciones infantiles en el campo de la investigación y cuenta usted con todo el apoyo de sus autoridades para participar en este grupo. Pero espera que el proyecto insista mucho en las necesidades y los problemas de

los jóvenes inmersos en la prostitución (los que tienen 16 años o menos). Si le pidieran que financiara la participación de los jóvenes en el proceso de investigación apoyaría tal solicitud en su organización (pero se plantearían problemas si pidieran a su organización que financiara a un grupo de jóvenes mayores, de 18 años o más). *Expondrá esta idea en un momento adecuado del debate.*

7. *Trabajador de un proyecto de la Organización Infantil Nacional B*

Su proyecto se ocupa de mujeres jóvenes, incluidas algunas que se han desvinculado de los servicios de atención. Ha participado usted en la creación de un proyecto de hogar seguro. Su organización sigue una política firme de igualdad de oportunidades, que incluye asegurar que en cualquier participación en una investigación se plantearán las cuestiones del género y la raza, una política que usted apoya totalmente. Le satisface «en principio» la disposición de los otros miembros del grupo a reconocer su postura, pero es consciente de que a algunos les puede molestar su insistencia.

Conoce a la persona de la Organización de Indigentes de Middlepoint y se entera de que va a proponer un *grupo consultivo de jóvenes* que va a trabajar como grupo centrado a lo largo del proyecto. Le encanta la idea, siempre que el grupo no lo formen únicamente chicos blancos.

8. *Representante de la Asociación de Responsables de Trabajo Social (ACOSW): Presidente de su Comité Infantil y Familiar*

Lleva usted algún tiempo defendiendo la política y la práctica de ACOSW en este campo y quiere que cuente con los recursos necesarios. Le preocupa que la investigación pueda verse perjudicada por personas que la vayan a utilizar para criticar el trabajo social, pero está satisfecho de las actitudes que muestran los componentes del grupo (aunque no está seguro de las posturas de otras personas de las organizaciones). Acaba de salir usted de un período de relaciones frías con algunas de las principales organizaciones infantiles, pero ha colaborado de forma provechosa con la policía en el desarrollo de una política conjunta sobre la prostitución infantil. Querrá asegurarse de que el grupo informe exhaustivamente a la policía sobre cualquier decisión que se tome en la reunión de hoy.

Lo que más le interesa es que la investigación produzca unos resultados que sirvan para configurar la política. Su objetivo principal

en esta reunión es *garantizar que se planifique adecuadamente la difusión final de la investigación*. Querrá usted que se adopten unas decisiones claras sobre las conferencias de prensa, las entrevistas en la radio y en la televisión que se hagan al final del estudio y sobre el tipo de publicaciones que se acuerde.

CAPÍTULO 8. REVISIÓN CULTURAL

En el capítulo 8 (páginas 211-212) describía e ilustraba una revisión cultural. Leamos de nuevo esa parte del capítulo, en especial el ejemplo. Imaginemos que planificamos una entrevista con una persona de las siguientes:

1. Una persona mayor que no quiere pensar en una residencia para la tercera edad.
 2. Una joven que dice que ha sufrido el acoso sexual de su profesor.
 3. Un joven de 16 años que debe comparecer ante el tribunal, acusado de exigir dinero con amenazas en su barrio.
 4. Una mujer que ha recaído en una depresión.
 5. Una mujer que tiene problemas para atender a su hijo de 25 años, que padece una grave discapacidad de aprendizaje y vive en casa.
- Hagamos una lista de los diferentes temas en los que deberíamos centrar nuestra autoindagación para una revisión personal y cultural.
 - Escribamos un párrafo sobre *dos* de los temas de la lista.

El ejercicio debe durar unos 20 o 30 minutos.

CAPÍTULO 9. ANÁLISIS DE LA ENTREVISTA EVALUATIVA

Consigamos la transcripción de una entrevista cualitativa que incluya una evaluación. A veces encontraremos fragmentos más extensos de transcripciones en investigaciones de evaluación cualitativa que se hayan publicado, por ejemplo en Pithouse y Atkinson (1988). Otra posibilidad es que el personal o los colegas de cualquier departamento de ciencias sociales de alguna universidad nos faciliten el acceso a un conjunto de transcripciones de entrevistas, garantizando siempre el anonimato.

Leamos de nuevo el apartado del capítulo 9 (páginas 243-247) en el que se habla del análisis de las entrevistas de evaluación, y tomemos unas breves notas sobre nuestra transcripción respecto a la *codificación indígena* y a los *vínculos narrativos*.

BIBLIOGRAFÍA

- Acker, J., K. Barry y J. Esseveld, «Objectivity and truth: problems in doing feminist research», *Women's Studies International Forum*, vol. 6, nº 4, 1983, págs. 423-435.
- Altheide, D. y J. Johnson, «Criteria for assessing interpretive validity in qualitative research», en N. Denzin e Y. Lincoln (comps.), *Handbook of Qualitative Research*, Thousand Oaks, Sage, 1994.
- , «Ethnography and justice», en G. Miller y R. Dingwall (comps.), *Context and Method in Qualitative Research*, Londres, Sage, 1997.
- Anderson, G., «Critical ethnography in education: origins, current status, and new directions», *Review of Educational Research*, vol. 59, nº 3, 1989, págs. 249-270.
- Atkinson, P., *Medical Talk*, Londres, Sage, 1995.
- Atkinson, P. y S. Delamont, «Bread and dreams or bread and circuses? A critique of case study research in evaluation», en Hammersley, M. (comp.), *Controversies in the Classroom*, Buckingham, Open University Press, 1993.
- Atkinson, P., S. Delamont y M. Hammersley, «Qualitative research traditions: a British response to Jacob», *Review of Educational Research*, 58, nº 2, 1988, págs. 231-250.
- Atkinson, R., *The Life Story Interview*, Thousand Oaks, Sage, 1998.
- Baker, C., «A second look at interviews with adolescents», *Journal of Youth and Adolescence*, vol. 12, nº 6, 1983, págs. 501-519.
- Barnes, M. y G. Wistow (comps.), *Researching User Involvement*, Leeds, Nuffield Institute for Health Services Studies, University of Leeds, 1992.
- , «Involving carers in planning and review», en A. Connor y S. Black (comps.), *Performance, Review and Quality in Social Care*, Londres, Jessica Kingsley, 1994.
- Barone, T. E., «Beyond theory and method: a case of critical story telling», *Theory into Practice*, vol. 31, nº 2, 1992, págs. 142-146.
- Bassey, M., «Does action research require sophisticated research methods?», en D. Hustler, A. Cassidy y E. Cuff (comps.), *Action Research in Classrooms and Schools*, Londres, Allen and Unwin, 1986.
- Becker, H., *Sociological Work*, Chicago, Aldine Publishing Company, 1970.

- , «Theory: the necessary evil», en D. Flinders y G. Mills (comps.), *Theory and Concepts in Qualitative Research: Perspectives from the Field*, Nueva York, Teachers College, Columbia University Press, 1993.
- Beresford, P., «Researching citizen involvement: a collaborative or colonising enterprise?», en M. Barnes y G. Wistow (comps.), *Researching User Involvement*, Leeds, Nuffield Institute for Health Services Studies, University of Leeds, 1992.
- Bernstein, R., *The Restructuring of Social and Political Theory*, Oxford, Blackwell, 1976.
- Best, J., *Images of Issues: Typifying Contemporary Social Problems*, Nueva York, Aldine de Gruyter, 1989.
- Bierman, K. y L. Schwartz, «Clinical child interviews: approaches and development consideration», *Child and Adolescent Psychiatry*, vol. 3, n° 4, 1986, págs. 309-325.
- Bloom, M., «Single-system evaluation», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Bloor, M., «On the routinised nature of work in people-processing agencies: the case of adeno-tonsillectomy assessments in ENT outpatient clinics», en A. Davis (comp.), *Relationships Between Doctors and Patients*, Farnborough, Saxon House, 1978.
- , «Notes on member validation», en R. Emerson (comp.), *Contemporary Trends in Field Research*, Boston, Little, Brown, 1983.
- , «Addressing social problems through qualitative research», en D. Silverman (comp.), *Qualitative Research: Theory, Method and Practice*, Londres, Sage, 1997a.
- , «Techniques of validation in qualitative research», en G. Miller y R. Dingwall (comps.), *Context and Method in Qualitative Research*, Londres, Sage, 1997b.
- Bloor, M. y N. McKeganey, «Outstanding practices: evaluative aspects of a descriptive study of eight therapeutic communities», *International Journal of Therapeutic Communities*, vol. 8, n° 3, 1987, págs. 273-285.
- , «Ethnography addressing the practitioner», en J. Gubrium y D. Silverman (comps.), *The Politics of Field Research: Sociology Beyond Enlightenment*, Newbury Park, Sage, 1989.
- Blumer, H., *Symbolic Interactionism: Perspective and Method*, Englewood Cliffs, NJ, Prentice-Hall, 1969 (trad. cast.: *Interaccionismo simbólico*, Barcelona, Hora, 1982).
- Bogdan, R. y S. Taylor, «A positive approach to qualitative evaluation and policy research in social work», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.
- Booth, D. (comp.), *Rethinking Social Development*, Harlow, Longman, 1994.

- Bowen, D., «The delights of learning to apply the life history method to school non-attenders», en B. Broad y C. Fletcher (comps.), *Practitioner Social Work Research in Action*, Londres, Whiting and Birch, 1993.
- Brody, S., *The Effectiveness of Sentencing*, Londres, Her Majesty's Stationery Office, 1976.
- Bulmer, M., *The Uses of Research*, Londres, Allen and Unwin, 1982.
- Bull, R. e I. Shaw, «Constructing causal accounts in social work», *Sociology*, vol. 26, n° 4, 1992, págs. 635-649.
- Burgess, R., «Some fieldwork problems in teacher-based research», *British Educational Research Journal*, vol. 6, n° 2, 1980, págs. 165-173.
- Burgess, R., C. Pole, K. Evans y C. Priestley, «Four studies from one or one study from four? Multi-site case study research», en A. Bryman y R. Burgess (comps.), *Analysing Qualitative Data*, Londres, Routledge, 1994.
- Butler, I. y H. Williamson, *Children Speak: Children, Trauma and Social Work*, Harlow, Longman, 1994.
- Campbell, D., «Distinguishing differences of perception from failures of communication in cross-cultural studies», en F. Northrop y H. Livingston (comps.), *Cross-Cultural Understanding: Epistemology in Anthropology*, Nueva York, Harper and Row, 1964.
- , «Qualitative knowing in action research», en M. Brenner y P. Marsh (comps.), *The Social Context of Methods*, Londres, Croom Helm, 1978.
- , «Degrees of freedom and the case study», en T. Cook y C. Reichardt (comps.), *Qualitative and Quantitative Methods in Evaluation Research*, Beverly Hills, Sage, 1979 (trad. cast.: *Métodos cualitativos y cuantitativos en investigación evaluativa*, Madrid, Morata, 2000).
- , «Foreword», en R. K. Yin, *Case Study Research*, Thousand Oaks, Sage, 1994.
- , «Can we overcome worldview incommensurability/relativity in trying to understand the other?», en R. Jessor, A. Colby y R. Shweder (comps.), *Ethnography and Human Development*, Chicago, University of Chicago Press, 1996.
- Campbell, D. y J. Stanley, *Experimental and Quasi-experimental Designs for Research*, Chicago, Rand McNally, 1966.
- Cancian, F. y C. Armstead, «Participatory research», en E. F. Borgatta y M. Borgatta (comps.), *Encyclopaedia of Sociology*, Nueva York, Macmillan, 1992.
- Carr, W. y S. Kemmis, *Becoming Critical: Education, Knowledge and Action Research*, Londres, Falmer Press, 1986 (trad. cast.: *Teoría crítica de la enseñanza: la investigación-acción en la formación del profesorado*, Barcelona, Martínez Roca, 1988).

- Chambers, R., *Rural Development: Putting the Last First*, Harlow, Longman, 1983.
- Cheetham, J., «The evaluation of social work: priorities, problems and possibilities», en J. Cheetham y M. Kazi (comps.), *The Working of Social Work*, Londres, Jessica Kingsley Publishers, 1998.
- Chelimsky, E., «Politics, policy and research synthesis», *Evaluation*, vol. 1, n° 1, 1995, págs. 97-104.
- , «Thoughts for a new Evaluation Society», *Evaluation*, vol. 3, n° 1, 1997, págs. 97-118.
- Cicourel, A., *Method and Measurement in Sociology*, Nueva York, Free Press, 1964 (trad. cast.: *El método y la medida en sociología*, Madrid, Editora Nacional, 1982).
- Clandinin, J. y R. M. Connelly, «Personal experience methods», en N. Denzin y Y. Lincoln (comps.), *Handbook of Qualitative Methods*, Thousand Oaks, Sage, 1994.
- Clapham, D., R. Means y M. Munro, «Housing, the life course and older people», en S. Arber y M. Evandrou (comps.), *Ageing, Independence and the Life Course*, Londres, Jessica Kingsley, 1993.
- Clarke, R. y D. Cornish, *The Controlled Trial in Institutional Research*, Londres, Her Majesty's Stationery Office, 1972.
- Clifford, D., «Critical life histories: key anti-oppressive research methods and processes», en B. Humphries y C. Truman (comps.), *Rethinking Social Research: Anti-Discriminatory Approaches in Research Methodology*, Aldershot, Avebury, 1994.
- Cochran-Smith, M. y S. Lytle, «Research on teaching and teacher research: the issues that divide», *Educational Researcher*, vol. 19, n° 2, 1990, págs. 2-22.
- Coffey, A. y P. Atkinson, *Making Sense of Qualitative Data*, Londres, Sage, 1996.
- Cohler, B., «The human services, the life story, and clinical research», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.
- Comstock, D., «A method for critical research», en F. Bredo y W. Feinburg (comps.), *Knowledge and Values in Educational Research*, Filadelfia, Temple University Press, 1982.
- Cook, T., «Postpositivist critical multiplism», en R. Shotland y M. Mark (comps.), *Social Science and Social Policy*, Beverly Hills, Sage, 1985.
- Cook, T. y D. Campbell, *Quasi-Experimentation*, Chicago, Rand McNally, 1979.
- Corsaro, W., *Friendship and Peer Culture in the Early Years*, Nueva Jersey, Ablex, Norwood, 1985.
- Crabtree, B. y W. Miller, *Doing Qualitative Research*, Newbury Park, Sage, 1992a.

- , «The analysis of narratives from a long interview», en M. Stewart, F. Tudiver, M. Bass, E. Dunn y P. Norton (comps.), *Tools for Primary Care Research*, Newbury Park, Sage, 1992b.
- Cronbach, L., *Designing Evaluations of Educational and Social Programs*, San Francisco, Jossey-Bass, 1982.
- , «Social inquiry by and for earthlings», en D. Fiske y R. Shweder (comps.), *Metatheory in Social Science: Pluralisms and Subjectivities*, Chicago, University of Chicago Press, 1986.
- Cronbach, L., S. Ambron, S. Dornbusch, R. Hess, R. Hornik, D. Phillips, D. Walker y S. Weiner, *Toward Reform of Program Evaluation*, San Francisco, Jossey-Bass, 1980.
- Davies, M. y E. Kelly, «The social worker, the client and the social anthropologist», *British Journal of Social Work*, vol. 6, n° 2, 1976, págs. 213-231.
- Davis, A., «Who needs user research? Service users as research subjects or participants», en M. Barnes y G. Wistow (comps.), *Researching User Involvement*, Leeds, Nuffield Institute for Health Services Studies, University of Leeds, 1992.
- Dean, R., «How good are you?», *Voluntary Housing*, enero de 1994.
- Denzin, N., *The Research Act*, Nueva York, McGraw Hill, 1989.
- Dullea, K. y A. Mullender, «Evaluation and empowerment», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Ebbutt, D., «Educational action research: some general concerns and specific quibbles», en R. G. Burgess (comp.), *Issues in Educational Research: Qualitative Methods*, Londres, Falmer Press, 1985.
- Edwards, M., «The irrelevance of development studies», *Third World Quarterly*, vol. 11, n° 1, 1989, págs. 116-135.
- , «Rethinking social development: the search for relevance», en D. Booth (comp.), *Rethinking Social Development*, Harlow, Longman, 1994.
- Eisner, E., «Educational connoisseurship and criticism: their form and functions in educational evaluation», en D. M. Fetterman (comp.), *Qualitative Approaches to Evaluation in Education*, Nueva York, Praeger, 1988.
- , «The meaning of alternative paradigms for practice», en E. Guba (comp.), *The Paradigm Dialog*, Newbury Park, Sage, 1990.
- , *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*, Nueva York, Macmillan, 1991a (trad. cast.: *El ojo ilustrado: indagación cualitativa y mejora de la práctica educativa*, Barcelona, Paidós, 1998).
- , «Taking a second look: educational connoisseurship revisited», en M. W. McLaughlin y D. Phillips (comps.), *Evaluation and Education at Quarter Century*, Chicago, University of Chicago Press, 1991b.

- Elden, M. y R. Chisholm, «Emerging varieties of action research», *Human Relation*, vol. 46, nº 2, 1993, págs. 121-141.
- Elks, M. y K. Kirkhart, «Evaluating effectiveness from the practitioner's perspective», *Social Work*, vol. 38, nº 5, 1993, págs. 554-563.
- Elliott, J., «Changing contexts of educational evaluation: the challenge for methodology», *Studies in Education*, nº 17, 1991, págs. 215-238.
- Ellsworth, E., «Why doesn't this feel empowering? Working through the myths of critical pedagogy», *Harvard Educational Review*, vol. 59, nº 3, 1989, págs. 297-324.
- Emerson, R. y M. Pollner, «On the uses of members' responses to researchers' accounts», *Human Organisation*, vol. 47, nº 3, 1988, págs. 189-198.
- England, H., *Social Work as Art*, Londres, Allen and Unwin, 1986.
- Epstein, L., «The trouble with the researcher-practitioner idea», *Social Work Research*, vol. 20, nº 2, 1996, págs. 113-118.
- Eraut, M., «Handling value issues», en C. Adelman (comp.), *The Politics and Ethics of Evaluation*, Londres, Croom Helm, 1984.
- Erikson, E., «The nature of clinical inference», en D. Lerner (comp.), *Evidence and Inference*, Illinois, Free Press, 1959.
- Everitt, A. y P. Hardiker, *Evaluating for Good Practice*, Londres, Macmillan, 1996.
- Everitt, A., P. Hardiker, J. Littlewood y A. Mullender, *Applied Research for Better Practice*, Londres, Macmillan, 1992.
- Filstead, W., «Qualitative methods – a needed perspective in evaluation research», en T. Cook y C. Reichardt (comps.), *Qualitative and Quantitative Methods in Evaluation Research*, Beverley Hills, Sage, 1979 (trad. cast.: *Métodos cualitativos y cuantitativos en investigación evaluativa*, Madrid, Morata, 2000).
- Finch, J., «Social policy and education: problems and possibilities of using qualitative research», en R. Burgess (comp.), *Issues in Educational Research: Qualitative Methods*, Londres, Falmer Press, 1985.
- , *Research and Policy: the Uses of Qualitative Methods in Social and Educational Research*, Londres, Falmer Press, 1986.
- Fine, G. y K. Sandstrom, *Knowing Children: Participant Observation with Minors*, Newbury Park, Sage, 1988.
- Firestone, W., «Alternative arguments for generalising from data as applied to qualitative research», en *Educational Researcher*, vol. 22, nº 4, 1993, págs. 16-23.
- Fisher, M., *Speaking of Clients*, Sheffield, University of Sheffield Joint Unit for Social Services Research, 1983.
- , «Users' experiences of agreements in social care», en M. Barnes y G. Wistow (comps.), *Researching User Involvement*, Leeds, Nuffield Institute for Health Services Studies, University of Leeds, 1992.

- , «Research, knowledge and practice in community care», *Issues in Social Work Education*, vol. 17, nº 2, 1997, págs. 17-30.
- Fisher, M., P. Marsh y D. Phillips, con E. Sainsbury, *In and Out of Care: The Experiences of Children, Parents and Social Workers*, Londres, Batsford, 1986.
- Fook, J. (comp.), *The Reflective Researcher: Social Workers' Theories of Practice Research*, St Leonards, NSW, Allen and Unwin, 1996.
- Fortune, A., «Ethnography in social work», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.
- Fournier, D. (comp.), *Reasoning in Evaluation: Inferential Links and Leaps*, San Francisco, American Evaluation Association/Jossey-Bass, 1995.
- Fournier, D. y N. Smith, «Clarifying the merits of argument in evaluation practice», *Evaluation and Program Planning*, vol. 16, nº 4, 1993, págs. 315-323.
- Frankland, J. y M. Bloor, «Some issues arising in the systematic analysis of focus group materials», en R. Barbour y J. Kitzinger (comps.), *Developing Focus Group Research: Politics, Theory and Practice*, Londres, Sage, 1999.
- Fuller, R. y A. Petch, *Practitioner Research: the Reflexive Social Worker*, Buckingham, Open University Press, 1995.
- Gage, N., «The paradigm wars and their aftermath», *Educational Researcher*, vol. 18, nº 7, 1989, págs. 4-10.
- Garrison, J., «The impossibility of atheoretical research», *Journal of Educational Thought*, vol. 22, nº 1, 1988, págs. 21-25.
- , «Realism, Dewey and pragmatism», *Educational Researcher*, vol. 21, nº 1, 1994, págs. 5-14.
- Geertz, C., *The Interpretation of Cultures*, Nueva York, Basic Books, 1973 (trad. cast.: *Interpretación de las culturas*, Barcelona, Gedisa, 1988).
- , *Local Knowledge*, Nueva York, Basic Books, 1983 (trad. cast.: *Conocimiento local*, Barcelona, Paidós, 1994).
- Gelsthorpe, L., «Response to Martin Hammersley's paper on feminist methodology», en *Sociology*, vol. 26, nº 2, 1992, págs. 213-218.
- Gibbs, A., «Focus Groups», *Social Research Update*, nº 19, Department of Sociology, University of Surrey, 1997, <<http://www.soc.surrey.ac.uk/sru/SRU19.html>>.
- Gibbs, L., E. Gambrill, J. Blakemore, A. Begun, A. Keniston, A. Peden y J. Lefcowitz, «A measure of critical thinking about practice», *Research on Social Work Practice*, vol. 5, nº 2, 1995, págs. 193-204.
- Goldstein, H., «Ethnography, critical inquiry and social work practice», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.

- Goode, D., «Kids, culture and innocents», *Human Relations*, vol. 9, n° 1, 1986, págs. 83-106.
- Gould, N., «Qualitative practice evaluation», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Gould, N. e I. Taylor (comps.), *Reflective Learning for Social Work*, Aldershot, Ashgate, 1996.
- Greene, J., «Three views on the nature and role of knowledge in social science», en E. Guba (comp.), *The Paradigm Dialog*, Newbury Park, Sage, 1990.
- , ««The practitioner's perspective», *Curriculum Inquiry*, 21, n° 1, 1992, págs. 39-45.
- , «The role of theory in qualitative program evaluation», en J. Flinders y G. Mills (comps.), *Theory and Concepts in Qualitative Research*, Nueva York, Teachers College Press, 1993.
- , «Qualitative program evaluation: practice and promise», en N. Denzin e Y. Lincoln (comps.), *Handbook of Qualitative Research*, Thousand Oaks, Sage, 1994.
- , «Qualitative evaluation and scientific citizenship», *Evaluation*, vol. 2, n° 3, 1996, págs. 277-289.
- Greene, J. y C. McClintock, «The evolution of evaluation methodology», *Theory into Practice*, vol. 30, n° 1, 1991, págs. 13-21.
- Guba, E., «The alternative paradigm dialog», en E. Guba (comp.), *The Paradigm Dialog*, Newbury Park, Sage, 1990.
- Guba, E. e Y. Lincoln, *Fourth Generation Evaluation*, Newbury Park, Sage, 1989.
- Hakim, C., *Research Design*, Londres, Allen and Unwin, 1988.
- Hall, B., «Participatory research: an approach for change», *Convergence*, vol. 8, n° 2, 1979, págs. 24-31.
- Hammersley, M., *What's Wrong With Ethnography?*, Londres, Routledge, 1992.
- , «On the teacher as researcher», *Educational Action Research*, vol. 1, n° 3, 1993a, págs. 425-445.
- , «On methodological purism», *British Educational Research Journal*, vol. 19, n° 4, 1993b, págs. 339-341.
- , «On practitioner ethnography», en M. Hammersley (comp.), *Controversies in Classroom Research*, Buckingham, Open University Press, 1993c.
- , «Ethnography, policy making and practice in education», en D. Halpin y B. Troyna (comps.), *Researching Education Policy: Ethical and Methodological Issues*, Londres, Falmer Press, 1994.
- , *The Politics of Social Research*, Londres, Sage, 1995.
- Hammersley, M. y R. Gomm, «Bias in social research», *Sociological Research Online*, vol. 2, n° 1, 1997, <<http://www.socresonline.org.uk/socresonline/2/1/2.html>>.

- Hart, E. y M. Bond, *Action Research for Health and Social Care*, Buckingham, Open University Press, 1995.
- Harvey, L., *Critical Social Research*, Londres, Unwin Hyman, 1990.
- Hawkesworth, M., «Knowers, knowing, known: feminist theory and claims of truth», *Signs: Journal of Women in Culture and Society*, vol. 14, n° 3, 1989, págs. 533-555.
- Heap, J., «Constructionism in the rhetoric and practice of fourth generation evaluation», *Evaluation and Program Planning*, vol. 18, n° 1, 1995, págs. 51-61.
- Hedrick, T., «The interaction of politics and evaluation», *Evaluation Practice*, vol. 9, n° 1, 1988, págs. 5-14.
- Heron, J., *Co-operative Inquiry: Research into the Human Condition*, Londres, Sage, 1996.
- Holman, B., «Research from the underside», *British Journal of Social Work*, vol. 17, n° 6, 1987, págs. 669-683.
- Holstein, J. y J. Gubrium, *The Active Interview*, Thousand Oaks, Sage, 1995.
- Hornsby-Smith, M., «Gaining access», en N. Gilbert (comp.), *Researching Social Life*, Londres, Sage, 1993.
- House, E., *Evaluating With Validity*, Beverly Hills, Sage, 1980.
- , «Evaluation and social justice: where are we now?», en M. McLaughlin y D. Phillips (comps.), *Evaluation and Education: at Quarter Century*, Chicago, Chicago University Press, 1991a.
- , «Realism in research», *Educational Researcher*, vol. 20, n° 6, 1991b, págs. 2-9.
- , *Professional Evaluation*, Newbury Park, Sage, 1993.
- , «Putting things together coherently: logic and justice», en D. Fournier (comp.), *Reasoning in Evaluation: Inferential Links and Leaps*, San Francisco, American Evaluation Association/Jossey-Bass, 1995.
- Humphrey, C. y K. Pease, «Effectiveness measurement in the Probation Service: a view from the troops», *Howard Journal*, vol. 31, n° 2, 1992, págs. 31-52.
- Humphreys, L., *Tearoom Trade*, Chicago, Aldine, 1975.
- Humphries, B., «From critical thought to emancipatory action: contradictory research goals?», *Sociological Research Online*, 1997, <<http://www.socresonline.org.uk/socresonline/2/1/3.html>>.
- , «Feminist evaluation», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Jacob, E., «Qualitative research traditions: a review», *Review of Educational Research*, vol. 57, n° 1, 1987, págs. 1-50.
- Jarrett, R., «Focus group interviewing with low income minority populations», en D. Morgan (comp.), *Successful Focus Groups: Advancing the State of the Art*, Newbury Park, Sage, 1993.

- Johnson, J., «In dispraise of justice», *Symbolic Interaction*, vol. 18, n° 2, 1995, págs. 191-205.
- Kelly, A., «Action research: what is it and what can it do?», en R. Burgess (comp.), *Issues in Educational Research*, Londres, Falmer Press, 1985.
- Kirkhart, K., «Seeking multicultural validity: a postcard from the road», *Evaluation Practice*, vol. 16, n° 1, 1995, págs. 1-12.
- Kirkpatrick, I. y M. Lucio (comps.), *The Politics of Quality*, Londres, Routledge, 1995.
- Kitzinger, J., «Focus groups: method or madness?», en M. Boulton (comp.), *Challenge and Innovation: Methodological Advances in Social Research on HIV/AIDS*, Londres, Taylor and Francis, 1994.
- Knodel, J., «The design and analysis of focus group studies: a practical approach», en D. Morgan (comp.), *Successful Focus Groups*, Newbury Park, Sage, 1993.
- Kreuger, R., *Focus Groups: a Practical Guide for Applied Research*, Thousand Oaks, Sage, 1994.
- Kushner, S., «Consumers and heroes: a critical review of some recent writings of Michael Scriven», *Evaluation*, vol. 5, n° 3, 1997, págs. 363-374.
- Lakatos, I., «Falsification and the methodology of scientific research programmes», en I. Lakatos y A. Musgrave (comps.), *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press, 1970.
- Lang, N., «Integrating the data processing of qualitative research and social work practice to advance the practitioner as knowledge builder: tools for knowing and doing», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.
- Lather, P., «Issues of validity in openly ideological research», *Interchange*, vol. 17, n° 4, 1986a, págs. 63-84.
- , «Research as praxis», *Harvard Educational Review*, vol. 56, n° 3, 1986b, págs. 257-277.
- , *Getting Smart: Feminist Research and Pedagogy with/in the Postmodern*, Nueva York, Routledge, 1991.
- , «Critical frames in educational research: feminist and post-structural perspectives», *Theory Into Practice*, vol. 31, n° 2, 1992, págs. 87-99.
- Lee, R., *Doing Research on Sensitive Topics*, Londres, Sage, 1993.
- Leplin, J. (comp.), *Scientific Realism*, Berkeley, University of California Press, 1984.
- Lerner, D. (comp.), *Evidence and Inference*, Nueva York, Free Press, 1959.
- Lewis, H., «Ethics and the managing of service effectiveness in social welfare», en R. Patti, J. Poertner y C. Rapp (comps.), *Managing for Service Effectiveness in Social Welfare Organisations*, Nueva York, Harworth, 1988.

- Lincoln, Y., «The making of a constructivist: a remembrance of transformations past», en E. Guba (comp.), *The Paradigm Dialog*, Newbury Park, Sage, 1990.
- Lincoln, Y. y E. Guba, «But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation», en D. D. Williams (comp.), *Naturalistic Evaluation*, San Francisco, Jossey-Bass, col. «New Directions in Program Evaluation», vol. 30, 1986a.
- , «Research, evaluation and policy analysis: heuristics and disciplined inquiry», *Policy Studies Review*, vol. 5, n° 3, 1986b, págs. 546-565.
- , «Ethics: the failure of positivist science», *Review of Higher Education*, vol. 12, n° 3, 1989, págs. 221-240.
- Lofland, J. y L. Lofland, *Analysing Social Settings*, Belmont, Wadsworth, 1995.
- Lyman, S. y M. Scott, *A Sociology of the Absurd*, Nueva York, Appleton-Century-Crofts, 1970.
- Mandell, N., «The least adult role in studying children», *Journal of Contemporary Ethnography*, vol. 16, n° 4, 1988, págs. 433-467.
- Martin, M., «Developing a feminist participative research framework», en B. Humphries y C. Truman (comps.), *Rethinking Social Research: Anti-Discriminatory Approaches in Research Methodology*, Aldershot, Avebury, 1994.
- , «Issues of power in the participatory research process», en K. de Koning y M. Martin (comps.), *Participatory Research in Health*, Londres, Zed Books, 1996.
- Martin, R., *Oral History in Social Work*, Thousand Oaks, Sage, 1995.
- , «Histories in social work», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Mayer, J. y N. Timms, *The Client Speaks*, Londres, Routledge and Kegan Paul, 1970.
- McCracken, G., *The Long Interview*, Newbury Park, Sage, 1988.
- McKeganey, N. y M. Bloor, «Spotting the invisible man: the influence of male gender on fieldwork relations», *British Journal of Sociology*, vol. 42, n° 2, 1991, págs. 195-210.
- McKeganey, N., I. MacPherson y D. Hunter, «How "they" decide: exploring professional decision making», *Research, Policy and Planning*, vol. 6, n° 1, 1988, págs. 15-19.
- McLennan, G., «Feminism, epistemology and postmodernism: reflections on current ambivalence», *Sociology*, 29, n° 2, 1995, págs. 391-409.
- McNamara, D., «The outsider's arrogance: the failure of participant observers to understand classroom events», *British Educational Research Journal*, vol. 6, n° 2, 1980, págs. 113-125.

- Miles, M. y A. Huberman, *Qualitative Data Analysis: an Expanded Sourcebook*, Thousand Oaks, Sage, 1994.
- Miller, J., *Creating Spaces and Finding Voices: Teachers Collaborating for Empowerment*, Nueva York, State University of New York Press, 1990.
- , «Exploring power and authority issues in a collaborative research project», *Theory into Practice*, vol. 21, n° 2, 1992, págs. 165-172.
- Miller, J. y M. Martens, «Hierarchy and imposition in collaborative inquiry: Teacher-researchers' reflections on recurrent dilemmas», *Educational Foundations*, vol. 4, n° 4, 1990, págs. 41-59.
- Miller, W. y B. Crabtree, «Depth interviewing: the long interview approach», en M. Stewart, F. Tudiver, M. Bass, E. Dunn y P. Norton (comps.), *Tools for Primary Care Research*, Newbury Park, Sage, 1992.
- , «Clinical research», en N. Denzin e Y. Lincoln (comps.), *Handbook of Qualitative Methods*, Thousand Oaks, Sage, 1994.
- Morgan, D. H. J., *The Family: Politics and Social Theory*, Londres, Routledge, 1985.
- Morgan, D. L. (comp.), *Successful Focus Groups: Advancing the State of the Art*, Newbury Park, Sage, 1993.
- , *Focus Groups as Qualitative Research*, Newbury Park, Sage, 1997.
- Oliver, M., «Changing the social relations of research production?», *Disability, Handicap and Society*, vol. 7, n° 2, 1992, págs. 101-114.
- Opie, I. y P. Opie, *The Lore and Language of Schoolchildren*, Oxford, Oxford University Press, 1959.
- , *Children's Games of Street and Playground*, Oxford, Oxford University Press, 1969.
- Outhwaite, W., *New Philosophies of Social Science*, Londres, Macmillan, 1987.
- Palumbo, D., «Politics and evaluation», en D. Palumbo (comp.), *The Politics of Program Evaluation*, Newbury Park, CA, Sage, 1987.
- Parker, W., «Interviewing children: problems and promise», *Journal of Negro Education*, vol. 53, n° 1, 1984, págs. 18-28.
- Parlett, M. y D. Hamilton, *Evaluation as Illumination*, Edimburgo, Centre for Research in the Educational Sciences, 1972.
- Patton, M., «The evaluator's responsibility for utilization», *Evaluation Practice*, vol. 9, n° 1, 1988a, págs. 5-24.
- , «Paradigms and pragmatism», en D. Fetterman (comp.), *Qualitative Approaches to Evaluation in Education*, Nueva York, Praeger, 1988b.
- , *Qualitative Evaluation*, Newbury Park, Sage, 1990.
- Pawson, R. y N. Tilley, *Realistic Evaluation*, Londres, Sage, 1997a.
- , «An introduction to scientific realist evaluation», en E. Chelmsky y W. Shadish (comps.), *Evaluation for the 21st Century*, Thousand Oaks, Sage, 1997b.

- Payne, G. y E. Cuff (comps.), *Doing Teaching: the Practical Management of Classrooms*, Londres, Batsford, 1982.
- Phillips, D., *Philosophy, Science and Social Inquiry*, Nueva York, Pergamon, 1987a.
- , «Validity in qualitative research», *Education and Urban Society*, vol. 20, n° 1, 1987b, págs. 9-24.
- , «Postpositivistic science: myths and realities», en E. Guba (comp.), *The Paradigm Dialog*, Newbury Park, Sage, 1990a.
- , «Subjectivity and objectivity: an objective inquiry», en E. Eisner y A. Peshkin (comps.), *Qualitative Inquiry in Education*, Nueva York, Teachers College, Columbia University, 1990b.
- Pithouse, A., *Social Work as an Invisible Trade*, Aldershot, Avebury, 1998.
- Pithouse, A. y P. Atkinson, «Telling the case: occupational narrative in a social work office», en N. Coupland (comp.), *Styles of Discourse*, Londres, Croom Helm, 1988.
- Pitman, M. y J. Maxwell, «Qualitative approaches to evaluation: models and methods», en M. LeCompte y J. Preissle (comps.), *The Handbook of Qualitative Research in Education*, San Diego, Academic Press, 1992.
- Plaut, T., S. Landis y J. Trevor, «Focus groups and community mobilisation», en D. L. Morgan (comp.), *Successful Focus Groups: Advancing the State of the Art*, Newbury Park, Sage, 1993.
- Popkewitz, T., «Whose future? Whose past?», en E. Guba (comp.), *The Paradigm Dialog*, Newbury Park, Sage, 1990.
- Popper, K., *The Open Society and Its Enemies*, vol. 2, Londres, Routledge and Kegan Paul, 1966 (trad. cast.: *La sociedad abierta y sus enemigos*, vol. 2, Barcelona, Paidós, 1984).
- , *Conjectures and Refutations: the Growth of Scientific Knowledge*, Londres, Routledge, 1989 (trad. cast.: *Conjeturas y refutaciones: el desarrollo del conocimiento científico*, Barcelona, Paidós, 1994).
- Pryce-Jones, M., «Critical incident technique as a method of assessing patient satisfaction», en R. Fitzpatrick y A. Hopkins (comps.), *Measurement of Patients' Satisfaction with their Care*, Londres, Royal College of Physicians, 1993.
- Pulice, R., «Qualitative evaluation methods in the public sector: understanding and working with constituency groups in the evaluation process», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.
- Reason, P. (comp.), *Participation in Human Inquiry*, Londres, Sage, 1994a.
- , «Three approaches to participative inquiry», en N. Denzin e Y. Lincoln (comps.), *Handbook of Qualitative Research*, Thousand Oaks, Sage, 1994b.

- Reason, P. y J. Rowan (comps.), *Human Inquiry: a Sourcebook of New Paradigm Research*, Chichester, John Wiley, 1981.
- Rees, S., «No more than contact», *British Journal of Social Work*, vol. 4, nº 3, 1974, págs. 255-279.
- Rees, S. y A. Wallace, *Verdicts on Social Work*, Londres, Edward Arnold, 1982.
- Reichardt, C. y T. Cook, «Beyond qualitative versus quantitative methods», en T. Cook y C. Reichardt (comps.), *Qualitative and Quantitative Methods in Evaluation Research*, Beverly Hills, Sage, 1979 (trad. cast.: *Métodos cualitativos y cuantitativos en investigación evaluativa*, Madrid, Morata, 2000).
- Reid, W., «Service effectiveness and the social agency», en R. Patti, J. Portner y C. Rapp (comps.), *Managing for Effectiveness in Social Welfare Organizations*, Nueva York, Haworth, 1988.
- , «Change-process research: a new paradigm?», en L. Videka-Sherman y W. Reid (comps.), *Advances in Clinical Social Work Research*, Silver Spring, MD, NASW Press, 1990.
- , «Reframing the epistemological debate», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994a.
- , «Field testing and data gathering on innovative practice interventions in early development», en J. Rothman y E. Thomas (comps.), *Intervention Research: Design and Development of Human Service*, Nueva York, Haworth Press, 1994b.
- , «The empirical practice movement», *Social Service Review*, vol. 68, nº 2, 1994c, págs. 165-184.
- , «Research on task-centred practice», *Social Work Research*, 21, nº 3, 1997, págs. 132-137.
- Reid, W. y P. Hanrahan, «Recent evaluations of social work: grounds for optimism», *Social Work*, nº 27, 1982, págs. 328-340.
- Reid, W. y A. Shyne, *Brief and Extended Casework*, Nueva York, Columbia University Press, 1969.
- Reid, W. y P. Zettergren, «A perspective on empirical practice», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Renzetti, C. y R. Lee (comps.), *Researching Sensitive Topics*, Newbury Park, Sage, 1992.
- Rist, R., «Influencing the policy process with qualitative research», en N. Denzin e Y. Lincoln (comps.), *Handbook of Qualitative Research*, Thousand Oaks, Sage, 1994.
- Rodwell, M. y D. Woody, «Constructivist evaluation: the policy/practice context», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.

- Rorty, R., *Philosophy and the Mirror of Nature*, Princeton, NJ, Princeton University Press, 1979 (trad. cast.: *La filosofía y el espejo de la naturaleza*, Madrid, Cátedra, 1989).
- , «Feminism and pragmatism», *Radical Philosophy*, n° 59, 1991, págs. 3-14.
- Rossi, P., «No good applied social research goes unpunished», *Society*, n° 25, noviembre-diciembre de 1987, págs. 74-79.
- Rothschild Report, *A Framework for Government Research and Development*, Cmnd 4814, Londres, Her Majesty's Stationery Office, 1971.
- Ruckdeschel, R., «Qualitative clinical research and evaluation: obstacles, opportunities and issues», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Ruckdeschel, R., P. Earnshaw y A. Firrek, «The qualitative case study and evaluation: Issues, methods and examples», en E. Sherman y W. Reid (comps.), *Qualitative Research in Social Work*, Nueva York, Columbia University Press, 1994.
- Sackett, D. y T. Rosenberg, «On the need for evidenced based medicine», *Health Economist*, vol. 4, n° 2, 1995, págs. 249-254.
- Schon, D., *The Reflective Practitioner: How Professionals Think in Action*, Nueva York, Basic Books, 1983 (trad. cast.: *El profesional reflexivo: cómo piensan los profesionales cuando actúan*, Barcelona, Paidós, 1998).
- , «The crisis of professional knowledge and the pursuit of an epistemology of practice», *Journal of Interprofessional Care*, vol. 6, n° 1, 1992, págs. 49-63.
- , «Reflective inquiry in social work practice», en P. Hess y E. Mullen (comps.), *Practitioner-Researcher Partnerships*, Washington, DC, NASW Press, 1995.
- Schwab, J., «The practical: a language for curriculum», *School Review*, noviembre de 1969, págs. 1-23.
- Schwandt, T., «Theory for the social sciences: crisis of identity and purpose», en D. Flinders y G. Mills (comps.), *Theory and Concepts in Qualitative Research: Perspectives from the Field*, Nueva York, Teachers College Press, 1993.
- , «Evaluation as practical hermeneutics», *Evaluation*, vol. 3, n° 1, 1997, págs. 69-83.
- Scott, D., «Practice wisdom: the neglected source of practice research», *Social Work*, vol. 35, n° 6, 1990, págs. 564-568.
- Scriven, M., «Objectivity and subjectivity in educational research», en L. Thomas (comp.), *Philosophical Redirection of Educational Research*, Washington, National Society for the Study of Education, 1972.
- , «New frontiers of evaluation», *Evaluation Practice*, vol. 7, n° 1, 1986, págs. 7-44.

- , «The logic of evaluation and evaluation practice», en D. Fournier (comp.), *Reasoning in Evaluation: Inferential Links and Leaps*, San Francisco, American Evaluation Association/Jossey-Bass, 1995.
- , «The theory behind practical evaluation», *Evaluation*, vol. 2, n° 4, 1996, págs. 393-404.
- , «Truth and objectivity in evaluation», en E. Chelimsky y W. Shadish (comps.), *Evaluation for the 21st Century*, Thousand Oaks, Sage, 1997.
- Searight, P., *Utilizing Qualitative Research to Self-Evaluate Direct Practice*, tesis doctoral, Saint Louis University, 1988.
- Selwyn, J., «Applying to adopt: the experience of rejection», *Adoption and Fostering*, vol. 15, n° 3, 1991, págs. 26-29.
- , «“Mirror, mirror on the wall. Who is the fairest of them all?” Involuntary childlessness and identity», en I. Butler e I. Shaw (comps.), *A Case of Neglect? Children's Experiences and the Sociology and Childhood*, Aldershot, Avebury, 1996.
- Shadish, W., «Philosophy of science and quantitative-qualitative debates: thirteen common errors», *Evaluation and Program Planning*, 18, n° 1, 1995, págs. 63-75.
- Shadish, W., T. Cook y L. Leviton, *Foundations of Program Evaluation: Theories of Practice*, Newbury Park, Sage, 1990.
- Shaw, I., *Evaluating in Practice*, Aldershot, Ashgate, 1996a.
- , «Unbroken voices: children, young people and qualitative methods», en I. Butler e I. Shaw (comps.), *A Case of Neglect? Children's Experiences and the Sociology of Childhood*, Aldershot, Avebury, 1996b.
- , *Be Your Own Evaluator: a Guide to Reflective and Enabling Evaluating*, Wrexham, Prospects Publishing, 1997.
- , «Practising evaluation», en J. Cheetham y M. Kazi (comps.), *The Working of Social Work*, Londres, Jessica Kingsley Publishers, 1998.
- , «Seeing the trees for the wood: the politics of evaluating in practice», en B. Broad (comp.), *The Politics of Research and Evaluation*, Birmingham, Venture Press, 1999.
- Shaw, I. y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Shaw, I. y A. Shaw, «Game plans, buzzes and sheer luck: doing well in social work», *Social Work Research*, vol. 21, n° 2, 1997a, págs. 69-79.
- , «Keeping social work honest: evaluating as profession and practice», *British Journal of Social Work*, vol. 27, n° 6, 1997b, págs. 847-869.
- Shaw, I., H. Williamson y N. Parry-Langdon, «Developing models for day services», *Social Policy and Administration*, vol. 26, n° 1, 1992, págs. 73-86.
- Sheldon, B. y G. Macdonald, «Implications for practice of recent social work effectiveness research», *Practice*, vol. 6, n° 3, 1989-1990, págs. 211-218.

- Sherman, E., «Change process research: no new paradigm but a necessary complement to the old paradigm», en L. Videka-Sherman y W. J. Reid (comps.), *Advances in Clinical Social Work Research*, Silver Spring, MD, NASW Press, 1990.
- Sherman, E. y W. Reid (comps.), *Qualitative Research and Social Work*, Nueva York, Columbia University Press, 1994.
- Siebert, M., «Adult elicited child behaviour: the paradox of measuring social competence through interviewing», en J. Cook-Gumperz, W. Corsaro y J. Streeck (comps.), *Children's Worlds and Children's Language*, Berlín, Mouton de Gruyter, 1986.
- Silverman, D., *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction*, Londres, Sage, 1993.
- , *Discourses of Counselling: HIV Counselling as Social Interaction*, Londres, Sage, 1997a.
- , «The logics of qualitative research», en G. Miller y R. Dingwall (comps.), *Context and Method in Qualitative Research*, Londres, Sage, 1997b.
- Simon, H., *Getting to Know Schools in a Democracy*, Londres, Falmer Press, 1987.
- Sinclair, I., *Hostels for Probationers*, Londres, Her Majesty's Stationery Office, 1971.
- Smith, J., «Interpretive inquiry: a practical and moral activity», *Theory into Practice*, vol. 31, n° 2, 1992, págs. 100-106.
- , «Hermeneutics and qualitative inquiry», en D. Flinders y G. Mills (comps.), *Theory and Concepts in Qualitative Research: Perspectives from the Field*, Nueva York, Teachers College Press, 1993.
- Smith, J. y L. Heshusius, «Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers», *Educational Researcher*, vol. 15, n° 1, 1986, págs. 4-12.
- Smith, L. y P. Klein, «Qualitative research and evaluation: triangulation and multi-methods reconsidered», en D. D. Williams (comp.), *Naturalistic Evaluation*, New Directions in Program Evaluation, n° 30, San Francisco, Jossey-Bass, 1986.
- Stacey, J., «Can there be a feminist ethnography?», *Women's Studies International Forum*, vol. 11, n° 1, 1988, págs. 21-27.
- Stake, R., «Retrospective on "The Countenance of Educational Evaluation"», en M. McLaughlin y D. Phillips (comps.), *Evaluation and Education at Quarter Century*, Chicago, University of Chicago Press, 1991.
- , «Case studies», en N. Denzin e Y. Lincoln (comps.), *Handbook of Qualitative Research*, Thousand Oaks, Sage, 1994.
- , *The Art of Case Study Research*, Thousand Oaks, Sage, 1995 (trad. cast.: *La investigación con estudio de casos*, Madrid, Morata, 1998).

- , «Advocacy in evaluation: a necessary evil?», en E. Chelimsky y W. Shadish (comps.), *Evaluation for the 21st Century*, Thousand Oaks, Sage, 1997.
- Stake, R. y D. Trumbull, «Naturalistic generalizations», *Review Journal of Philosophy and Social Science*, n° 7, 1982, págs. 1-12.
- Stenhouse, L., «The teacher as researcher», en M. Hammersley (comp.), *Controversies in Classroom Research*, Buckingham, Open University Press, 1993.
- Stufflebeam, D., «Professional standards and ethics for evaluators», en M. McLaughlin y D. Phillips (comps.), *Evaluation and Education at Quarter Century*, Chicago, University of Chicago Press, 1991.
- Swigonski, M., «Feminist standpoint theory and questions of social work research», *Affilia*, vol. 8, n° 2, 1993, págs. 171-183.
- Tanner, K. y P. Le Riche, «Work in progress: the contribution of observation to the development of good practice and evaluation», en I. Shaw y J. Lishman (comps.), *Evaluation and Social Work Practice*, Londres, Sage, 1999.
- Thomas, E., «Research and service in single-case experimentation: conflicts and choices», *Social Work Research and Abstracts*, vol. 14, n° 4, 1978, págs. 309-325.
- Timms, N., *The Language of Social Casework*, Londres, Routledge and Kegan Paul, 1968.
- Tizard, J., I. Sinclair y R. Clarke (comps.), *Varieties of Residential Experience*, Londres, Routledge and Kegan Paul, 1975.
- Toulmin, S., *The Uses of Argument*, Cambridge, Cambridge University Press, 1958.
- Trend, M. G., «On the reconciliation of qualitative and quantitative analyses», en T. Cook y C. Reichardt (comps.), *Qualitative and Quantitative Methods in Evaluation Research*, Beverly Hills, Sage, 1979 (trad. cast.: *Métodos cualitativos y cuantitativos en investigación evaluativa*, Madrid, Morata, 2000).
- Tripp, D. H., «Co-authorship and negotiation: the interview as an act of creation», *Interchange*, vol. 14, n° 3, 1983, págs. 32-45.
- Turner, M. y W. Zimmerman, «Acting for the sake of research», en J. Wholey, H. Hatry y K. Newcomer (comps.), *Handbook of Practical Program Evaluation*, San Francisco, Jossey-Bass, 1994.
- U. S. Department of Education, *What Works: Research about Teaching and Learning*, Washington, DC, US Department of Education, 1986.
- Vanderplaat, M., «Beyond technique: issues in evaluating for empowerment», *Evaluation*, vol. 1, n° 1, 1995, págs. 81-96.
- Warwick, D., «Tearoom Trade: means and ends in social research», en M. Bulmer (comp.), *Social Research Ethics*, Londres, Macmillan, 1982.

- Wasoff, F. y R. Dobash, «Simulated clients in “natural” settings: constructing a client to study professional practice», *Sociology*, vol. 26, n° 2, 1992, págs. 333-349.
- , *The Simulated Client: a Method for Studying Professionals Working with Clients*, Aldershot, Avebury, 1996.
- Weiss, C., «Knowledge creep and decision accretion», *Knowledge, Creation, Diffusion, Utilisation*, vol. 1, n° 3, 1980, págs. 381-404.
- , «The stakeholder approach to evaluation: origins and promise», en A. Bryk (comp.), *Stakeholder-Based Evaluation*, San Francisco, Jossey-Bass, 1983a.
- , «Toward the future of stakeholder approaches in evaluation», en A. Bryk (comp.), *Stakeholder-Based Evaluation*, San Francisco, Jossey-Bass, 1983b.
- , «Where politics and evaluation meet», en D. Palumbo (comp.), *The Politics of Program Evaluation*, Newbury Park, Sage, 1987.
- , «If programme decisions hinged only on information», *Evaluation Practice*, vol. 9, n° 1, 1988, págs. 15-28.
- , *Evaluation: Methods for Studying Policies and Programs*, Nueva York, Prentice-Hall, 1998.
- Whitmore, E., «Empowerment in program evaluation: a case example», *Canadian Social Work Review*, vol. 7, n° 2, 1990, págs. 215-229.
- , «To tell the truth: working with oppressed groups in participatory approaches to inquiry», en P. Reason (comp.), *Participation in Human Inquiry*, Londres, Sage, 1994.
- Wholey, J., «Assessing the feasibility and likely usefulness of evaluation», en J. Wholey, H. Hatry y K. Newcomer (comps.), *Handbook of Practical Program Evaluation*, San Francisco, Jossey-Bass, 1994.
- Williams, B., «Patient satisfaction: a valid concept?», *Social Science Medicine*, vol. 38, n° 4, 1994, págs. 509-516.
- Williams, D. D., «When is naturalistic evaluation appropriate?», en D. D. Williams (comp.), *Naturalistic Evaluation*, San Francisco, Jossey-Bass, col. «New Directions in Program Evaluation», vol. 30, 1986.
- Williams, R., *Keywords: a Vocabulary of Culture and Society*, Londres, Fontana, 1976.
- Willis, P., *Learning to Labour: How Working Class Kids Get Working Class Jobs*, Farnborough, Saxon House, 1977 (trad. cast.: *Aprendiendo a trabajar*, Madrid, Akal, 1988).
- Witkin, S. y S. Gottschalk, «Alternative criteria for theory evaluation», *Social Service Review*, vol. 62, n° 2, 1988, págs. 211-224.
- Woods, R. y A. Pollard, *Sociology and Teaching: a New Challenge for the Sociology of Education*, Londres, Croom Helm, 1988.

- Woolgar, S. y D. Pawluch, «Ontological gerrymandering: the anatomy of social problems explanations», *Social Problems*, vol. 32, n° 3, 1985, págs. 214-227.
- Yin, R. K., *Case Study Research*, Thousand Oaks, Sage, 1994.
- Zeller, R., «Focus group research in sensitive topics: setting the agenda without setting the agenda», en D. L. Morgan (comp.), *Successful Focus Groups: Advancing the State of the Art*, Newbury Park, Sage, 1993.

ÍNDICE DE NOMBRES

- Abse, D., 250
- Acker, J., 57, 59, 167, 235, 249-250
- Altheide, D., 106, 151, 187-188
- Atkinson, P., 50, 85, 94, 160, 173, 195-196, 263-264
- Atkinson, R., 218
- Baker, C., 215
- Barnes, M., 202-203
- Barone, T., 217
- Bassey, M., 155
- Becker, H., 23, 42, 82, 100, 120
- Beresford, P., 201
- Best, J., 105-106
- Bierman, K., 213
- Bloom, M., 119
- Bloor, M., 110, 172-174, 228, 257-258
- Blumer, H., 87
- Bogdan, R., 130, 200, 205
- Bowen, D., 217
- Bull, R., 134-135
- Bulmer, M., 142
- Burgess, R., 159, 198, 226
- Butler, I., 214
- Campbell, D., 25, 29, 39-43, 87, 93, 96, 118, 129, 189-190, 259
- Carr, W., 57, 78, 97, 132-133, 158, 201-202
- Cheetham, J., 158
- Chelimsky, E., 17, 20, 22, 82, 107, 114, 124, 133, 137-141, 260-261
- Cicourel, A., 213
- Clandinin, J., 80, 219-220, 250
- Clapham, D., 218
- Cochran-Smith, M., 149.
- Coffey, A., 235
- Cohler, B., 217
- Comstock, D., 217
- Cook, T., 41, 82, 86-87, 118
- Corsaro, W., 216
- Crabtree, B., 155-156, 205, 210-211, 243-244, 246, 261, 268
- Cronbach, L., 15, 19, 39, 45-47, 69, 93, 97-99, 109, 119, 123, 125, 129-130, 133-134, 137-138, 145, 193-194, 196, 242, 253, 269
- Davies, M., 227
- Davis, A., 203
- Delamont, S., 50, 85, 94, 160, 195-196
- Denzin, N., 157, 259
- Dewey, J., 61, 64, 84
- Edwards, M., 57, 165-166, 168
- Eisner, E., 19, 30-32, 40, 62-65, 82-83, 114, 150, 169, 184-186, 262-263
- Elden, M., 201
- Elks, M., 170
- Elliott, J., 26
- Emerson, R., 257-258
- England, H., 169
- Epstein, L., 156-157
- Eraut, M., 111, 134, 241, 270
- Erikson, E., 97, 151, 171-172
- Everitt, A., 57
- Feyerabend, P., 260
- Filstead, W., 73, 86
- Finch, J., 20, 51, 106, 136, 143-144, 184, 197, 243, 251-252, 268
- Fine, G., 216

- Firestone, W., 264
 Fisher, M., 113, 203, 229
 Fook, J., 173
 Fortune, A., 149
 Fournier, D., 45, 178, 240-242, 245
 Freire, P., 77

 Gage, N., 83
 Garrison, C., 95, 117
 Geertz, C., 30-33, 62, 96-97, 103, 157, 177, 264-265
 Gibbs, A., 222
 Gibbs, L., 168
 Goffman, I., 264
 Goldstein, H., 168
 Goode, D., 214
 Gould, N., 119, 173
 Greene, J., 21, 29, 31, 64, 84, 86, 90, 94-95, 106, 113, 125, 131-132, 136, 188, 194, 270
 Guba, E., 18, 26-28, 65-67, 73-75, 77, 79-80, 86, 88-92, 103-104, 110, 114-117, 120, 178
 Gubrium, J., 205, 209-210, 245-246

 Hakim, C., 195-196
 Hammersley, M., 50, 74, 76, 82, 85, 87, 89, 97, 101-103, 107, 117, 145, 153, 160-161, 173, 248, 258, 266
 Hanrahan, P., 60-61
 Hardiker, P., 57
 Harvey, L., 57-58
 Hawkesworth, M., 152
 Heap, J., 67, 80, 91-92, 105
 Hedrick, T., 141
 Heron, J., 161-164, 234, 248-249
 Holman, R., 165
 Holstein, J., 205, 209-210, 244-246, 266
 Hornsby-Smith, N., 225
 House, E., 25-26, 43, 54-57, 63, 88, 114, 181, 183, 186-187, 192-194, 240, 261

 Huberman, A., 30-33, 75, 82, 189, 193, 253-254, 256
 Humphrey, C., 170
 Humphreys, L., 114-115
 Humphries, B., 78

 Jacob, E., 73-74
 Johnson, J., 106, 151, 187-188

 Kelly, A., 83, 94, 155
 Kemmis, S., 57, 78, 97, 132, 158, 201-202
 Kirkhart, K., 106, 170
 Kirkpatrick, I., 71
 Kitzinger, J., 221-222
 Knodel, J., 221
 Kuhn, T., 41, 73
 Kushner, S., 45, 49-50

 Lakatos, I., 76-77
 Lang, N., 157, 147-148
 Lather, P., 47, 57-59, 77, 103, 107-108, 248-249, 259
 Leplin, J., 86, 88
 Lerner, D., 151
 Lincoln, Y., 18, 26-28, 65-67, 79, 81, 103-104, 110, 114-117, 120, 178
 Lofland, J., 188, 193, 225-226, 234, 239, 253

 Macdonald, G., 18
 Mandell, N., 216
 Martin, M., 167
 Martin, R., 216
 McCracken, G., 131, 209-211, 243-244
 McKeganey, N., 174, 207-208, 228,
 McLennan, G., 80-81
 McNamara, D., 148, 173
 Miles, M., 30-33, 75, 82, 189, 193, 253-254, 256
 Miller, J., 57, 167, 232-234
 Miller, W., 155-156, 205, 210-211, 243-244, 246, 261, 268

- Morgan, D. L., 221
Morgan, D., 218
- Oliver, M., 57
Opie, I., 215
- Palumbo, D., 18
Parlett, M., 154
Patton, M., 52, 82-83, 139, 205
Pawson, R., 193
Phillips, D., 74-77, 88, 258-259
Pithouse, A., 173, 287
Pitman, M., 64, 67, 85
Popkewitz, T., 58, 77-78, 81
Popper, K., 41, 73, 76-77
Pryce-Jones, M., 230
Pulice, R., 182
- Rawls, J., 250
Reason, P., 77, 103, 105, 162-163
Rees, S., 213, 230
Reichardt, C., 86
Reid, W., 40-41, 59-62, 83-84, 88, 156, 191, 243
Rist, R., 142
Rodwell, M., 183
Rorty, R., 61, 84-85
Rossi, P., 270
Ruckdeschel, R., 157
- Sackett, D., 71
Schon, D., 120, 169, 171-173
Schwab, J., 97, 152
Schwandt, T., 67, 79, 84, 97, 152
Scott, J., 152, 219
Scriven, M., 18, 39-40, 43-45, 82-83, 92, 178, 226, 266
Searight, P., 231-232
Selwyn, J., 218
Shadish, W., 42-43, 45-46, 48, 50, 53, 82, 111, 117-119, 137, 178, 253
Shaw, A., 96, 135, 170-171, 277
Shaw, I., 23, 75, 77, 96, 134-135, 158, 170-171, 213, 230, 254, 256, 277
Sheldon, B., 18
- Sherman, E., 62, 156, 242
Siegert, M., 215
Silverman, D., 30, 173, 192, 239, 242-243, 253
Simon, H., 48, 56, 106, 155, 194-195, 199, 267-268
Sinclair, I., 199
Smith, J., 79-81, 89, 104-105, 117-118, 259
Smith, L., 150
Stacey, J., 116, 167, 184, 235
Stake, R., 15, 19, 40-41, 48-50, 64-65, 147, 151-152, 165, 168, 195-196, 261
Stenhouse, L., 154
Swigonski, M., 77
- Tanner, K., 215
Thomas, E., 120
Timms, N., 96, 153
Tizard, J., 199, 263
Toulmin, S., 241
Trend, M., 33-34, 252
Tripp, D., 248-249, 266-267
Trumbull, D., 49, 147, 151
Turner, M., 223
- Vanderplaat, M., 207
Warwick, D., 114-115
Wasoff, F., 223
Weiss, C., 19, 50-53, 71, 112-113, 174, 181, 206, 276
Whitmore, E., 57, 126-127, 167
Wholey, J., 124-126, 180-181, 205
Williams, B., 229
Williams, D., 33, 179
Williams, R.
Witkin, S., 25
Woods, R., 206
Woolgar, S., 90-92, 105
- Yin, R., 178, 195-199, 225
Zettergen, P., 61-62, 84

ÍNDICE ANALÍTICO

- Acceso, 225-227
American Evaluation Association, 16, 19
Análisis causal, 21, 33, 42-43, 46, 57, 78, 129, 189, 192-194, 243, 253-254
Asistencia social, 59-62, 135, 156-157
Autoría, 266-267
- Centre for Applied Research in Education (CARE), 49, 153
Codificación, 244-245
Confidencialidad, *véase* Ética
Conocimiento de sentido común, 29, 42, 93, 95-96
Conocimiento tácito, 151, 171
Consentimiento informado, *véase* Ética
Contar historias, 80, 156, 210, 217, 219-220, 268-269. *Véase también* Narrativa, relato
- Desarrollo mediante la evaluación, 59-60
Difusión, 267
Documentos, 219
- Educación, 19, 62-65, 153
Enseñanza, *véase* Educación
Entrevistas, 207-215, 222, 224, 243-246, 277-278, 287-288
Especialización y crítica, 63-64, 262
Estudio de casos, 46, 48-49, 157, 189, 192, 194-200, 263
Ética, 35, 55, 114-117, 183, 185-186, 234-237, 250-251
- Etnografía, 20, 149-150, 168, 173-174, 177, 190, 195, 264, 266
Evaluación, propósitos, 20-23, 33-34, 42-45, 101-102, 106-107, 123, 150-151, 177-178, 205, 239-240
afirmaciones de la evaluación, 240-242
análisis, 239-240, 242-244, 246-249, 252-254, 256, 260, 263-264, 268
de la evaluación cualitativa, 15, 29-34, 67-68, 143-144, 251
diseño, 177-178, 180-181, 183, 186, 189-190, 194-195, 197, 204
e investigación, 18, 20-21, 24-28
trabajo de campo, 205-207, 216, 221, 225, 234, 236
- Evaluación constructivista, 65-67, 79-80, 105-107, 183, 187-188, 210
Evaluación crítica, 57-58, 77-78, 132-133, 158, 207. *Véase también* Evaluación e investigación de defensa
Evaluación de la práctica, 119-120, 147-152, 155-158, 160-161, 163, 165-166, 168-169, 171-175, 232, 234, 247, 250
Evaluación de programas, 19, 124-129, 131, 134-136
Evaluación democrática, 26, 106, 160, 199, 267
Evaluación e investigación de defensa, 20, 49, 78, 107, 136-138, 165-169, 217-218, 233-235, 270

- Evaluación e investigación feministas, 166-167, 235, 249
 como postura teórica, 58-59, 152
 posmodernidad feminista, 80
 y pragmatismo, 84-86
- Evaluación e investigación participativas, 44, 49-50, 127, 161-168, 201-202, 204, 206, 221-222, 247-249, 266-268
- Evaluación emancipadora, *véase* Evaluación e investigación de defensa
- Evaluación formativa, 43, 45
- Evaluación iluminativa, 154
- Evaluación potenciadora, *véase* Evaluación e investigación de defensa
- Evaluación receptiva, 48, 50
- Evaluación reformista, 53-55, 106, 145, 187, 270
- Evaluación sin metas, 43-44, 92
- Evaluar en la práctica, 166-170
- Experimentos y cuasi-experimentos, 41-42, 93, 129
- Falsa conciencia, 79, 108, 174, 203
- Fundamentismo, 76
- Generalización, 48-50, 63, 65, 97, 109, 197-198, 215, 261-265
- Grupos centrados, 221-222
- Grupos de interés, 48
- Historias de la vida, 216-217
- Ilustración, *véase* Usos de la evaluación
- Indigencia, 236-237
- Instrumentalismo, *véase* Usos de la evaluación
- Intimidad *véase* Ética
- Investigación clínica, 155-156
- Investigación en la acción, 201-203
- Investigación sobre el desarrollo del Tercer Mundo*, 165
- Investigación y evaluación, 15, 28-30, 32, 64, 178
- Investigación y evaluación cuantitativas, 33-34, 55, 63-64, 98, 143-144, 260
- Investigación y evaluación de políticas, 50-52, 136-138, 140-144
- Ironía, 253
- Justicia, 54-57, 106-107, 186-188, 250-251
- Metaanálisis, 43, 260
- Metodología, 22-23, 53, 58, 117-120, 155, 160, 177-178, 205-207
- Narrativa, relato, 210, 244, 246. *Véase también* Contar historias
- Neomarxismo, 56
- Niños, 213-216
- Objetividad, 44-45, 92, 104-105
- Observación, 148, 207-209, 215-216, 231-232, 273-274
- Paradigmas, 40, 56, 66-68, 72-74, 79-83, 85-86, 113, 115-116, 118, 275
- Pluralismo, 46-47, 53-55, 67, 138
- Política, 50-51, 112-113, 140-141, 181
- Positivismo, 66, 75, 77, 115-116, 169
- Pospositivismo, 66, 75-77, 79, 116
- Práctica basada en pruebas, 15, 17-18, 71-72
- Práctica reflexiva, 62, 168-169, 171-172
- Pragmatismo, 53-54, 61, 81-86, 158
- Preguntas sobre satisfacción, 228-231
- Profesor como investigador, 149, 153, 158-159, 167, 226-227, 234
- Razonamiento práctico, 95, 97, 152. *Véase también* Conocimiento de sentido común
- Realismo, 41, 56, 73-75, 77, 86-88, 192-193

- Relativismo, 42, 65-66, 74, 88-90, 109, 259
- Relevancia, 47, 92-98, 165-166, 265-267, 269
- Representación de roles, juegos de simulación, 223-224, 279-287
- Responsabilidad, rendición de cuentas, 17, 21-22, 47, 130, 133-135
- Resultados, 16, 59-61, 130, 177-178, 180-181, 188-189, 191-194
- Revisión cultural, 131, 211, 287
- Rigor, 92-94
- Sabiduría práctica, 96, 152
- Sentimentalismo*, 23
- Simulaciones, 223-224
- Síntesis, *véase* Metaanálisis
- Sociología del curso de la vida, 217-218
- Temas sensibles, 227
- Teoría, 39-40, 76-77, 80-81, 93-95, 117, 119, 131-133, 144, 157, 160, 202-203, 264
- Triangulación, 108, 259
- Usos de la evaluación, 18, 53, 112-114, 119, 123, 136-137, 139, 141-142, 174, 276
- Validación de los miembros, *véase* Validez
- Validez, 100-104, 106-110, 256, 258, 260
 - externa, 46, 98, 109, 129, 189
 - interna, 41, 45-46, 189
 - validación de los miembros, 110, 256-257
- Valores, 78, 104, 110-111
- Viñetas, 223

La evaluación cualitativa
Ian F. Shaw

Paidós
Básica

Esta obra de Ian Shaw constituye una excelente introducción práctica a las estrategias y los métodos de la evaluación cualitativa. Ofrece la primera demostración coherente de cómo se pueden utilizar los planteamientos cualitativos en la práctica evaluadora cotidiana para superar las limitaciones de los enfoques cuantitativos más tradicionales.

La evaluación cualitativa abarca una amplia gama de temas: los fundamentos de la evaluación y sus tendencias actuales; los principales teóricos de la evaluación; la evaluación de programas, de políticas y de la práctica profesional; el diseño, la recogida de datos y el análisis de la evaluación. Además, en todos los capítulos se dan ejemplos para mostrar cómo aplicar los temas, los enfoques y los métodos que esos casos ilustran.

Se trata, en fin, de una perfecta introducción para estudiantes, investigadores y usuarios de la evaluación, especialmente en los campos de la educación, el trabajo social y otras profesiones afines. Un libro fundamental tanto para el sector público como para los organismos sin ánimo de lucro, así como para los investigadores cualitativos en el campo de todas las ciencias sociales.

Ian Shaw es profesor de Trabajo Social en la Universidad de Cardiff y, junto con Joyce Lishman, compilador de *Evaluation and Social Work Practice*.

ISBN 84-493-1426-7



9 788449 314261

Diseño: Mario Eskenazi